THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. II 1950 Part 2

THE EFFECT OF ONE EXPERIENCE UPON THE RECALL OF ANOTHER

BY

D. RUSSELL DAVIS and DURGANAND SINHA

(*From the Psychological Laboratory, University of Cambridge*)

An exploratory experiment was made to show how one experience may exert an influence upon the recall of another, and how both may lose their separate identities in memory, to become merged into what Bartlett has described as an active organization of related experiences. The original material was a story, and the experience interpolated between it and its recall was a picture which illustrated part of it, in some respects correctly, in others incorrectly. Some subjects were asked to recall the story, others the picture, after from one to four weeks, or in a few cases a year.

The picture was viewed under the influence of the attitudes induced by the story. These determined its selection from a number of other unrelated pictures, to be perceived as a related experience. In consequence only few details gained attention, and these were interpreted in terms of the attitudes induced by the story, in some cases contrary to the meaning given to them by the rest of the picture. Other details without significance in the theme of the story were neglected. In its turn, seeing the picture favoured the recall of some aspects of the theme of the story, details not reinforced by the picture tending to be left out in recall. Many picture details *intruded* into the reproductions of the story. This tendency was progressive, and at the end of a year picture and story details were not separated. Especially, the picture brought about changes in the points of emphasis in the story.

I

Introduction

The essence of Bartlett's theory of remembering is the idea that past experience operates as "an organized mass rather than as a group of elements each of which retains its specific character" (Bartlett, 1932). Experiences are retained in memory in the form of active organizations or schemata, a schema being a reification of experience conceived of as a plastic model which is reshaped by every new experience related to it. Schemata may be partially independent of one another, but tend to overlap and fuse.

The several experiments which he describes justify Bartlett's view of memories as constructions of related experiences undergoing continuous modification. Yet it cannot be denied that the memories of some experiences remain discrete, or at least relatively independent of one another, fixed and uncontaminated, as, e.g., the theory of Freud requires; for even if the clinical evidence is discounted—and this would hardly be reasonable—there are in everyone's personal experience impressive examples of events, usually of some special significance, which have resisted assimilation into general experience, and some aspects of which remain capable of relatively accurate or "literal" recall even after the passage of many years.

The theories of Bartlett and Freud are not incompatible with one another in this respect. Indeed, although this point has often been missed, Freud (e.g., 1892, pp. 29–34) emphasized the special character of the memories which give rise to hysterical phenomena. Thus he remarked upon their "wonderful freshness," "astonishing integrity and intensity of feeling" and "their exceptional fate in reference to all the ordinary processes of effacement," and suggested that they are "preserved with such freshness and affective force because the normal process of absorption by abreaction and by reproduction in a state of unrestrained association is denied them." On the other hand, he argued, the more usual process is for a memory to become "merged into the great complex of associations," and then to become "ranged alongside of other experiences which perhaps contradict it; thus it undergoes correction by means of other ideas."

It is pertinent therefore to enquire further into the conditions which favour the preservation in memory of the separate identity of an experience, or its merging into an "active organization" or "complex of associations" of related experiences. Before the attempt is made to overcome the difficulties of a direct experimental approach, however, the problem requires clearer definition, and in our opinion there is a need for a number of exploratory experiments. Accordingly, we have made several experiments to test the influence of interpolated material of various kinds upon the reproduction of an original experience. The present paper describes a study of this influence when the original material was a story and the interpolated material a picture which illustrated part of the story (though not being in complete agreement in every detail). Some subjects were asked to recall the story, and others the picture. The investigation was intended to throw light on the manner of recall and the kind of items recalled.

II

Material and Procedure

The story and picture were so chosen that the picture was relevant to, and could be expected to exert a big influence upon the recall of the story. The picture was a coloured postcard reproduction of Pieter Breughel's painting, "The Village Wedding," with which, although it is famous, our subjects were not familiar. Around the scene there depicted was written in a pompous style a story of about 750 words, partially descriptive and partially episodic. This story (which is printed as an appendix) gave to the scene of conviviality a significance quite different from that which the painter intended; its first part described a feud between two families and their reconciliation in the betrothal of the son of one with the daughter of the other; its latter part, a wedding feast, some details from the picture being rendered correctly, some being hinted at, some being deliberately misrepresented and others being left out; in this part of the story mention was made of several untoward incidents which nearly led to a revival of the feud.

The subjects were university students. They were divided into four groups, who were treated according to the plan shown in Table I. Of subordinate interest was the

TABLE I

Plan of the Experiment

	1st day, original experience	*4th day interpolated experience*	*4th to 28th day recall of*
Group A (21)	Reading of story	Presentation of picture	Story
Group B (28)	Reading of story	—	Story
Group C (7)	Reading of story	Presentation of picture	Picture
Group D (7)	—	Presentation of picture	Picture

The number of subjects is given in brackets.

comparison of the recall of the picture in Groups C and D, which showed how the story had influenced the perception and recall of the picture. Comparison of the recall by Groups A and B showed how the picture thus perceived then influenced the recall of the story.

In some respect this plan was similar to that of the experiments on the so-called retroactive inhibition of one habit by another. Our main concern was with the influence of the interpolated material upon the recall of the original material, and this may be thought of as akin to the influence of the performance of one habit upon the subsequent performance of another, when the two habits are similar in some respects. Then facilitation (positive transfer) or interference (negative transfer) is looked for. Only one overall measure of the performance has been obtained in the majority of experiments on this topic. In our experiment, however, the recall of picture or story was studied in a more qualitative than quantitative manner, to show in what respects there was facilitation and in what respects interference, and, furthermore, to what degree there was intrusion of the components of the one into the other.

The procedure was rather more complicated than the foregoing description indicates, for "The Village Wedding" was presented with six other unrelated pictures as part of a quasi-recognition test. The reason for this was twofold: to disguise the real purpose of the experiment; to obtain collateral information about the interaction of picture and story. Three or four days after the story was read before Groups A, B and C, each subject of Groups A and C was shown the seven postcards and asked to identify the one which depicted a scene described in the story. All the postcards were reproductions of paintings in colour. Although several of the other postcards showed rural scenes of festivities, every subject picked out "The Village Wedding," as had been expected. He was then given brief opportunity to examine it. The subjects of Groups A and B were then asked to recall the story on one occasion only, but at intervals varying from immediately to more than three weeks. Those of Group C were asked to recall the picture at similar intervals. Those of Group D, before whom the story was not read, were shown "The Village Wedding" by itself and were then treated in the same way as were those of Group C. Throughout, the instructions to the subjects were framed in such a way that, although the immediate demands were clear, there were no grounds for anticipating what would be required of them on later occasions.

III

Results

(i) *Perception of the Picture.*

Before coming to the effects of the picture on the recall of the story, it is necessary to comment upon the effects of the story on the perception and recall of the picture.

The story seemed to create certain expectations which served as a frame of reference for the perception of the picture. Thus when presented with the seven postcards, most subjects looked for a scene of a certain kind only; one subject said, for instance, "I was looking for a feast in a rustic setting." They had a general impression of what the picture was going to be like, and on that basis "The Village Wedding" was often picked out at once by some process of immediate matching; it was recognized at once as being relevant to the story and as complying with expectations. Similarly, several of the postcards were rejected immediately as "irrelevant" or as not having the right "atmosphere." A reproduction of a painting by Toulouse-Lautrec, for instance, was usually rejected quickly, even scornfully, as "too modern" or as "obviously French." Of interest was this reason for rejecting the same picture, "the dress and atmosphere do not fit," for nothing in the story defined the dress directly, although all would agree that the dress in Toulouse-Lautrec's picture would be out of keeping with such a peasant affair as the story described. Sometimes one other postcard was picked out for more careful consideration because it showed the right kind of scene; this was Pieter Breughel's "The Peasant Dance." It was then discarded, either because some detail was

incompatible with the story, e.g. the scene is an out-of-doors one, or because "The Village Wedding" was preferred, this picture containing details which had been mentioned in the story.

"The Village Wedding" having been chosen in this way, on the basis of general impression, subjects then usually confirmed their decision by looking for details mentioned in the story and expressed their confidence as soon as they had found them. Three details especially served this purpose: "pitchers of wine being filled," "crossed sheaves of corn" and "two servants bearing ash staves." These details were mentioned in the story, the latter two having a special significance in its theme. The former two were fairly prominent in the picture. With the last there commonly occurred a false perception; the two servants bearing ash staves being identified in the picture with what the subjects who had not heard the story described more correctly as "musicians" or sometimes as "bag-pipers."

In the majority of subjects, general impression played the primary, and details a secondary rôle in the selection of "The Village Wedding." Details determined the immediate choice in rather less than one-third of our subjects. Even these subjects, however, who employed a distinctly analytic method discarded other pictures on a basis of general impression. Some subjects expected to see certain details, which, as it happened, were absent from "The Village Wedding" and had imagined a scene with different detail; nevertheless they chose "The Village Wedding" and felt little doubt of the correctness of the choice, although they might express annoyance at the discrepancies between expectation and actuality.

Recall of the Picture.—The subjects who had heard the story tended to recall those details in the picture which had played an important rôle in the story, and which were easily located in the picture. On the other hand, details equally prominent in the picture, but not mentioned in the story, were not recalled at all or, if they were recalled, were treated as accessory. Many of these subjects clearly indicated in their reports that once "The Village Wedding" had been chosen, and the choice verified by one or more details, they had hardly bothered to examine the picture carefully. They interpreted the few details of the picture, which gained their attention, ascribing to them qualities derived from the story, e.g. a bridegroom was often identified, "sitting in the place of honour" and "appearing sad and downcast"; and they recalled those details for which the story had provided an interpretation.

Those who had not heard the story tended merely to enumerate the details of the picture, and their reports revealed careful and minute observation of detail. The number of subjects in Groups C and D was too small for quantitative comparisons to yield significant results, but there was little doubt that, considered from the point of view of the number of items recalled, those who had heard the story gave poorer accounts of the picture than those who had not heard the story; i.e., in respect of the number of items, hearing the story was a handicap in the recall of the picture.

(ii) *Recall of the Story.*

(*a*) *Amount of Recall.*—In order to obtain a quantitative estimate of the amount of recall, the original story was divided up arbitrarily into about 120 items (as has been done in the appendix), and a similar enumeration of items was carried out for each reproduction. Each item in a reproduction was examined with reference to the original story and then classified. The method was crude, but was regarded as sufficient for its purpose. It was found that the reproductions of those to whom the picture had been shown (Group A) contained more items than did those of the

other subjects (Group B). The size of the difference between Groups A and B is shown by the facts presented in Table II. The story was divided into two parts: the antecedent part which comprised the incidents leading up to the wedding feast;

TABLE II

MEDIAN NUMBER OF ITEMS RECALLED IN EACH SUB-GROUP—WHOLE STORY

	Recall after less than 8 days	*Recall after 21–28 days*
Group A	50	43
Group B	45	24
For the τ correlation, P is ..	0·02	0·005

The method of Whitfield (1947) was used for the calculation of probability, and median scores are cited here, for this method allowed the combination of several sub-groups differing in the interval before recall.

and the main part, the description of the feast itself. As might be expected, the interpolated experience favoured the recall of the main rather than the antecedent part (Table III).

TABLE III

MEDIAN NUMBER OF ITEMS RECALLED IN EACH SUB-GROUP—(1) ANTECEDENT AND (2) MAIN PARTS SEPARATELY

	Recall after less than 8 days		*Recall after 21–28 days*	
	(1)	(2)	(1)	(2)
Group A	15	35	14	29
Group B	16	29	8	16
For the τ correlation, P is ..	—	0·01	—	0·001

(*b*) *Importations.*—The items in each reproduction were divided into two classes: accurately recalled, simply transformed or distorted details, with counterparts in the original; importations, without counterparts in the original. A large part, but not the whole of the difference between Groups A and B in the number of items recalled was found to be due to the larger number of importations in Group A.

TABLE IV

MEDIAN NUMBER OF IMPORTATIONS IN EACH SUB-GROUP

	Recall after less than 8 days	*Recall after 21–28 days*	*For the τ correlation, interval x number of importations, P is*
Group A	4	7	0·01
Group B	3	4	0·05
For the τ correlation, group x number of importations, P is ..	—	0·01	—

Of these importations the majority were *intrusions*, i.e. details derived from the picture. Intrusions are one of three types of importation, the other two types being inventions and rationalizations. The three types have been treated together in Table IV, from which it will be seen that:

(i) In recall after less than 8 days the difference between Groups A and B is not significant.

(ii) In both groups there were more importations in recall after 21 days than after less than 8 days.

(iii) In recall after 21 days there were more importations in Group A than B.

These results are in agreement with the finding of Bartlett that, "the tendency to invent or to import new material from a different setting may increase considerably with lapse of time."

(*c*) *Mode of Recall.*—The recall of those who had not seen the picture tended to be of the *sequential* type; i.e. they reported merely a chain of events with a minimum of circumstantial detail. Also, they tended to use condensed or compressed forms of expression; for instance, in describing the main theme of the story, they used such expressions as "regaled with truly magnificent feast," "there was much feasting and merrymaking" and "the scene was a merry one." In contrast, the recall of those who had seen the picture was fuller and more graphic, e.g. "there were two long tables (*intrusion*); at one sat the two families, at the other the visitors (*invention*)": "Plates upon plates of food were brought (*accurate recall*)": "The servants stood oafishly eyeing the good things before them (*intrusion and distortion*)": "with stupid faces and mouths oafishly hanging down (*distortion*)."

In commenting upon their recall of the story, subjects reported what they had considered to be the crucial details, and from what, in the manner described by Bartlett, they had reconstructed the story. Those who had not seen the picture tended to mention as dominant details items from the antecedent part, e.g. the family feud, the straying of the sheep and the death of the father. On the other hand, dominant details were drawn more often from the main part of the story by those who had seen the picture, e.g. the crossed sheaves of corn and the entry of servants bearing staves. Thus the interpolated experience caused a redistribution and reshuffling of the points of emphasis in recall.

(*d*) *Fate of Certain Details.*—Two details especially were important to the theme of the story and were also prominent in the picture; these were: "crossed sheaves of corn" and "two servants bearing ash staves." The subjects who had seen the picture recalled the former almost invariably, and the latter frequently, but these details tended to drop out of the recall of the other subjects, being recalled after 21 days by less than half and less than one-fifth of them respectively. If they were recalled by these subjects, they were often distorted; e.g. "crossed sheaves of corn" were recalled as "a family crest" and on one occasion even as "on the wall there hung a sheep which everyone suspected was the stolen sheep." Thus some salient details were accentuated by the interpolated experience.

Other circumstantial details, such as "pitchers being filled with wine," which were often remarked upon in the picture, and which contributed in the story only to the atmosphere of feasting and gaiety, were included only occasionally in reproductions of the story. Even then they were remembered as picture details.

Recall After One Year.—A year after the reading of the story the nine subjects of Group A and the seven of Group B who were still available were asked to recall

as much as they could of the story. They were instructed to report what they recalled as it came along, and not to think it over and then present it in a coherent form.

In the number of items recalled there was no longer a difference which could be regarded as significant, but the preferential recall of primary details was still to be noted in those who had seen the picture; for instance, "crossed sheaves of corn" were recalled by all but one of these subjects, but were omitted altogether or grossly confused by the other subjects. These tended first to recall the "long-standing family feud," in such terms as "a common tale of family feud and killing in a primitive society" or as "a Romeo and Juliet story," and this part of the theme was the basis on which the rest of the story was reconstructed. On the other hand, in the recall of those who had seen the picture the main stress fell on the wedding feast, and nearly all of them referred to the picture as a first and dominant association; as one subject said, "the picture created a sort of setting and was the starting-point for the remembering of the story."

Most said that they could remember the picture better than the story. Generally subjects could not decide whether a detail which had been recalled came from the picture or the story, and mistakes were numerous, picture details appearing as intrusions in the recall of the story, e.g. "labourers with coloured hats"; "image of servants carrying plates on a tray"; "sheaves hanging on the wall over a table." These examples illustrate a confusion of picture and story details which was typical.

IV

Discussion

This experiment provided a realistic demonstration of the influence of one experience upon the recall of another. The story was written and presented in a manner intended to give it emphasis and character and to aid its distinction from experiences not under experimental control. In other respects, as has been said, story and picture were deliberately chosen to favour their merging with one another in memory, and several devices in the framing of the instructions to the subjects were employed for this purpose. As would be expected, these intentions were not realised completely. Although this aspect of the results has not been dealt with, the story did not escape the influence of the stream of experiences which preceded and succeeded it, and into which the experiment was inserted, for the recall of the subjects of both groups, A and B, showed much transformation and rationalization from this cause. Again, the picture and story were not entirely assimilated into one another, and each retained some degree of separate identity.

Now it can readily be demonstrated that under conditions other than those which obtained for our subjects, the two experiences can be perceived, retained and recalled without contamination of each other. The conditions of our experiment, however, caused the postcards to be viewed with the expectation that at least one of them would be related to the story. As a result six postcards were looked at cursorily and rejected as unrelated; they exerted no detectable influence upon the subsequent recall of the story. "The Village Wedding," however, was recognized as relevant to the story and was examined under the influence of the attitudes induced by it. Being evoked, these attitudes became modified by the picture; and thus modified, they governed the subsequent recall of both picture and story.

The evolution of attitudes in such processes as these may perhaps be conceived of as akin to the shaping of a skilled response by every new situation which evokes it.

If there is indeed a parallelism between the evolution of the attitudes governing perception and recall and the development of a skill, it is not surprising that neither the original attitudes nor the experiences can be recalled unmodified by subsequent experiences. At any rate, neither picture nor story could be recalled in our experiment, even after the shorter intervals, in anything like the forms which they would have taken, had they been independent of one another. No theory of remembering would be satisfactory, therefore, which treated them as independent experiences with a separate identity in memory. In the terms of Bartlett's theory, however, they may be held to have been retained in memory, each as a part of an actively-organized setting, the schemata representing them becoming closely connected to one another and compounded to a degree increasing with passage of time. The progressiveness of the fusion of the experiences argues against any theory which treats the reorganization of memories as similar to the reorganization of a configuration with the addition or substraction of an element. Evidence for this progressiveness is found in the comparison of recall after less than 8 days with that after 21–28 days, and with that after a year, when it seemed that picture and story had become inextricably mixed with one another.

Comments may now be made upon the nature of the interaction between the two experiences. Undoubtedly the effect of the story on the perception and recall of the picture was due to its verbal character as well as to its precedence in time. "Vocalization," Bartlett states, "favours the classification of presented material according to some rule or *idée directrice*." Our findings showed this tendency to be an important one. "The Village Wedding" was selected because its features complied with the general impressions which the story had created. So far as they can be conveyed in a few words, these impressions were: feud, smouldering enmity, uneasy truce, conviviality, in a rustic, possibly Flemish setting. "The Village Wedding" served as a representative of this general class of scenes, being identified as the correct representative because of the congruence of certain details, such as the crossed sheaves of corn. The story thus provided modes of abstraction, or principles of organization, which caused the features of the picture to be interpreted in terms of concepts of feud, disorderliness, conviviality, and the like. These concepts were not available to the subjects of Group B, who enumerated the particulars of the picture without generalization.

Another aspect of this effect of story on picture is of interest. Whereas the scene portrayed by "The Village Wedding" is usually regarded as one of gaiety, the story suggested it to be one of restlessness and suppressed excitement. Subjects could not rid themselves of this impression, although they might be aware of its origin and falseness. This tendentious attitude was stabilized and perpetuated to some degree by misinterpretations for which it was itself responsible; for instance, innocent and harmless musicians were perceived as servants threatening with weapons, and the throng at the back of the room was taken to be an unruly one. Probably misinterpretations of this kind have counterparts in everyday life, e.g. in rumours and paranoid ideas, when they corroborate or even enhance attitudes which have been adopted as the result of an experience of some special significance. Nevertheless, although sustained to some degree by misinterpretations, the impressions of feud and disorderliness were weakened by the picture, for those who had seen it emphasised conviviality, rather than feud and enmity, in their subsequent recall.

In its turn the picture had thus a considerable effect upon the recall of the story. Not only did it bring about a change in emphasis, but it also facilitated the recall of some features, led to the neglect of others and brought about intrusion of its

details. Possibly some of these results came about the more readily because it favoured the use of visual imagery rather than language; to this tendency may have been due the large number of intrusions. But the most striking effect (although one not easily quantified) was the change in emphasis. This was shown in a marked form in the recall of the story after a year; in the one group the story was reconstructed around the concept of a feud in a peasant society, in the other around the fragmentary recall of a wedding feast.

In conclusion, we may return once more to the question asked in the first section of the paper: what conditions are necessary to bring about the processes of merging of experiences which we have now described? Although it proved no hypothesis, the experiment suggested that the essential condition was the re-arousal by the interpolated material of the memories of the original material and of their attendant attitudes. This being done, a secondary condition was the congruence of some salient details.

Appendix

(The story, showing the arbitrary divisions into items)

"For many generations, | the peace | of the little village of Neuf | had been disturbed | by the feud | between the Weydens and the Loons. | Ancient grudges, | constantly revived, | had kept the two families at loggerheads, | and the slightest cause of friction | had been sufficient to fan a smouldering jealousy into a flame of murderous strife. | One day, however, | an aimless sheep | straying from Martin Weyden's field | into that of Henrik Loon | became the cause of a violent quarrel, | in which even the village folk were divided into two camps, | and which ended in the death of Henrik Loon, | the head of the Loon family. | Shortly before his father's death, Henrik Loon's heir | began secretly to court Hetta Weyden, | the daughter of his father's murderer, | and, becoming the head of the family, | hastened to declare his wish to abandon former rivalries and to cement a new cordiality between the two families | by marriage. | Martin Weyden accepted the gesture with generosity | and promised a sincere welcome to his new son-in-law. | The village folk were eager to forget their partisanship | and soon showed themselves willing to vie with one another | only in the elaborateness of their preparations for the wedding ceremony. |

The wedding feast, which was held at the Weyden farm, | was honoured even by the sheriff, whose efforts had often been necessary to curb the violence of their quarrels | and to arrange a truce between the families. | Sitting in a high-backed chair | at the head of the broad and laden table, | a dark Holbein cap | almost hiding his white hair | and the velvet of his nobility being somewhat incongruous | in the humble interior of the farmhouse, | he gazed anxiously | at Hans Loon, | the uncle of the bridegroom and the brother of the murdered man. | His solemn expression was unsuited | to the rustic boisterousness of the village folk. | But he had reasons to be alarmed, | for several unhappy incidents | made him doubt the permanence of the sudden friendship between the two families. | The bridegroom | and his relatives | had been greeted with the graciousness and the lavishness which the occasion demanded, | but in a prominent position | over the bridegroom's seat | there hung | the crossed sheaves of corn, | which were not only the traditional symbol of fertility in that agricultural community, | but were also the emblem of the Weyden family. | It could hardly have escaped the notice | of the Loons | that the rival emblem was thus flaunted; | and, on the other side, it was only too obvious | that the bridegroom's coat was made, | as all knew, | from the skin of the sheep | whose straying had led to the death of Henrik Loon. | Upon the bridegroom's head | was

ONR BRANCH LIBRARY
NAVAL RESEARCH LABORATORY

the blue cap | which had been worn last by Henrik Loon on the day of his death. | Indeed, ugly memories | seemed to haunt the young man, | as he sat looking sad and downcast | and hardly interested in the gay chatter of his bride.

"But nothing seemed to daunt the vivacious gossiping of the women, | and as dish after dish was brought | for the guests' pleasure, | and the coloured earthenware pitchers were filled | again and again | with sweet, white wine, | the danger that rancour would flare up again | seemed to grow less. | But just as the men's uneasiness was giving way to a careless enjoyment | of the many delicacies spread before them, | Hans Loon's two servants | entered the farmhouse, | bearing the two ash staves | which they had brought with them to the gathering. | Hans Loon had bidden them stand outside | and guard the horses and saddlebags of his family. | But, seeing the freedom with which the village folk took part in the celebrations, | the two oafish peasants | followed the stream of villagers into the farmhouse. | Although their arms threatened, | their faces showed only a yearning to partake of the rich food | which was being enjoyed by the villagers and the guests alike. | Their countenances were childlike in their stupidity, | as unable to restrain their appetite saliva dribbled from their mouths. | Yet Martin Weyden, angered by their rudeness | and by their display of arms, | called loudly upon Hans Loon | to dismiss them. | But Hans Loon had fallen asleep, | his burly figure slumping forward | and his long white beard falling upon the table. | His sword dangled by his side, | and his faithful hound | peered from under the table | at the strange human ceremony | where the tragedies of enmity were forgotten, | and where fears were lost | in the unrestrained joy of the wedding feast.|"

V

References

BARTLETT, F. C. (1932). *Remembering*. Cambridge.

FREUD, S., with BREUER, J. (1892). In: *Collected Papers*. Vol. I. London.

WHITFIELD, J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, **34**, 292–296.

(*Received 2nd January*, 1950.)

THE RELATIONSHIP BETWEEN AREA AND THRESHOLD IN THE CENTRAL FOVEA FOR LIGHTS OF DIFFERENT COLOURS

BY

E. N. WILLMER

(*From the Physiology Laboratory, Cambridge*)

The relationship between the size of the field and its threshold intensity has been plotted for small fields fixated on the central fovea. Within the limits investigated, no significant difference has been found whether this relationship has been determined for red or for violet light. While it is possible that there is a continuous relationship between the threshold intensity and the size of the field, in practice there is a rather sharp change in the relationship when the field size subtends about 5′ or 6′. Between this size and a field subtending about 20′, the expression $\sqrt{A}\, I = C$ describes the findings, but with smaller fields the relationship $A\, I = C$ is much more nearly, though not absolutely, true. Experimental difficulties may limit the extent to which the observations can be described by this second equation.

I

Introduction

It has been known for many years that an inverse relationship exists between the area of the retina stimulated and the brightness of the stimulating light which is necessary to evoke a sensation. For small areas, and particularly in the fovea, Ricco's law is approximately true and states that for threshold excitation the area times the intensity is constant. For larger areas and in the periphery of the retina, Piper's law is more nearly applicable. This law states that, at threshold, the square root of the area times the intensity is constant. Neither of these laws, however, exactly fits the facts, and Wald (1938) has introduced an expression, $(A - n_t)^k I = C$, which satisfied his data and other data for all but very large and very small fields. In this expression, A = area stimulated, n_t = a constant threshold number of elements, I = intensity and k and C are constants. This author came to the conclusion that two separate laws (Ricco's and Piper's) were not necessary, but that the data could be described by this one equation. Graham, Brown and Mote (1939) showed that any equation of the form $A^k I = C$ was unsatisfactory and produced another quantitative account of the area-intensity relationship which held good in the fovea up to a limit of about 1°, at which point rods probably began to vitiate the results. In the periphery a similar relationship was established which held good for field sizes between the limits of 10′ and about 10°, after which the threshold became independent of area. In order to establish whether this peripheral relationship was different for rods and cones Graham and Bartlett (1939) worked with both red and violet light in the periphery. They came to the conclusion that since the curves expressing their results for the area-intensity relationship for the two colours were parallel, there was no evidence that anything but rods contributed to the determination of the threshold. Baumgardt (1947), however, came to somewhat different conclusions. Arguing from van der Velden's experimental finding that two quanta absorbed within a certain time interval are sufficient to produce excitation, he was led to predict that Ricco's law ($A\, I = C$) should hold for areas of the retina

occupied by one independent or quasi-independent unit, i.e. receptor cell or more probably ganglion cell, and that Piper's law ($\sqrt{A}\ I = C$) should hold for areas containing more than one such unit. His experiments initiated by this hypothesis showed that with a blue light at a position 15° temporally from the fovea, the relationship $A\ I = C$ held for rod vision for areas between 2′ 12″ to 31′ 36″, and thus the quasi-independent unit appeared to be large and presumably might correspond to the ganglion cell which is known to spread its dendrites far and wide and to cover a large number of rods. On the other hand, when working with extreme red light and finding no photochromatic interval, so that he was probably stimulating cones, he found that Piper's law ($\sqrt{A}\ I = C$) was obeyed between 12′ 30″ and 31′ 36″, indicating the presence of many quasi-independent units within this area and suggesting that the ganglion cells activated by cones occupy smaller areas, which is of course in accordance with the histological findings (Polyak, 1941). More recently (Baumgardt, 1948) he has extended the observations to the fovea and has shown that under several different conditions with respect to duration of the flash and to pupil size there was no significant difference between the behaviour of the fovea in red light or green light. For example, he found that when no artificial pupil was used, the relationship $A^{0.95}\ I = C$ was obeyed within the limits 1·12′ and 6·72′ diameter of field. Between 6·72′ and 33·6′ the relationship was $A^{0.42}\ I = C$, thus showing that neither Ricco's nor Piper's law was exactly valid but they are each nearly so for the field size to which they are appropriate and that there is a change of law at about 6·72′.

Now, if Baumgardt is correct in supposing that the range within which Ricco's law holds good gives a measure of the size of the quasi-independent unit, then the foveal unit must correspond in size to the retinal field covered by an object subtending an angle of 6·72′. This is considerably above the size of field corresponding to a single cone but might relate again to the area served by a single ganglion cell or by certain types of bipolar cell, although the data for the visual acuity of the fovea would indicate the probability that the effective unit is considerably smaller than this. If, however, the size of the quasi-independent unit is the determining factor in bringing about the change of law, then it is relevant to enquire whether this unit differs according to the wavelength in a manner which might suggest different pathways for different colours. Since the central fovea behaves as though it had only two independent types of receptor and two corresponding pathways (König, 1894; Willmer, 1944) it is probable that one of these will be stimulated more intensely by light from the blue end of the spectrum while the other will respond more in the red region. It therefore seemed possible that some information might be obtained as to the size of the quasi-independent unit for each of these pathways by making measurements of the area-intensity relationship for the central fovea in red light and in violet light. As already mentioned, Baumgardt (1948) has found that there was no significant difference when red and green lights were used, but it might be that these two wavelengths are not far enough apart to stimulate the two receptor systems sufficiently exclusively to make any significant difference. This consideration together with the fact that there was a divergence between Baumgardt's observations with red and blue light in the periphery and the earlier observations of Graham and Bartlett, made it desirable to re-investigate the area-intensity relationship in the central fovea and to use violet and red test lights instead of green and red. By this means it might be possible to determine the size of the quasi-independent units in the fovea and to decide whether or not the units corresponding to the two receptors differed significantly in size.

II

Methods

The apparatus already described (Willmer, 1949) was used to illuminate two small fields with light of any colour or intensity required. The right-hand field subtended about 5′ at the eye and was illuminated throughout the experiments by light from an Ilford spectrum red filter (608) at an intensity about five times the threshold for the foveal centre of the subject. This field acted as a pilot light or fixation point and the subject was required to fixate on the left-hand edge of this field. By means of stops, the test field, which could be illuminated by either extreme red (Ilford 609) or violet (Ilford 601) was located with its centre 20′ away from the centre of the pilot field and horizontally to the left. The stops varied in size in such a way that, at the eye, their diameters subtended angles varying between 0·72′ and 19·4′. With accurate fixation these fields would all fall on the so-called rod-free area. Fixation was assisted by the usual dental impression technique for maintaining a constant head position. An artificial pupil, 2 mm. in diameter was used throughout the experiments. The subject could vary the intensity of the light in the left-hand field by turning a knob and he was allowed as much time as he liked to reach a decision as to what was the threshold position when he was certain of his central fixation. In this way the technique differed to some extent from that adopted by previous investigators. Between each reading the position of the lamp on the scale was altered, thus changing the light intensity. This was done by the operator so that the subject could have no idea from the position of the lamp or of any other part of the apparatus whether he was near threshold or not and the complete independence of each reading was thus secured. The fields of different sizes were given in random order, and the readings were repeated on several different occasions, so that any point in the diagrams is always based on the mean of at least four entirely independent readings and most of them are based on ten or more. The readings have all been made on two subjects whose vision and colour-vision were normal.

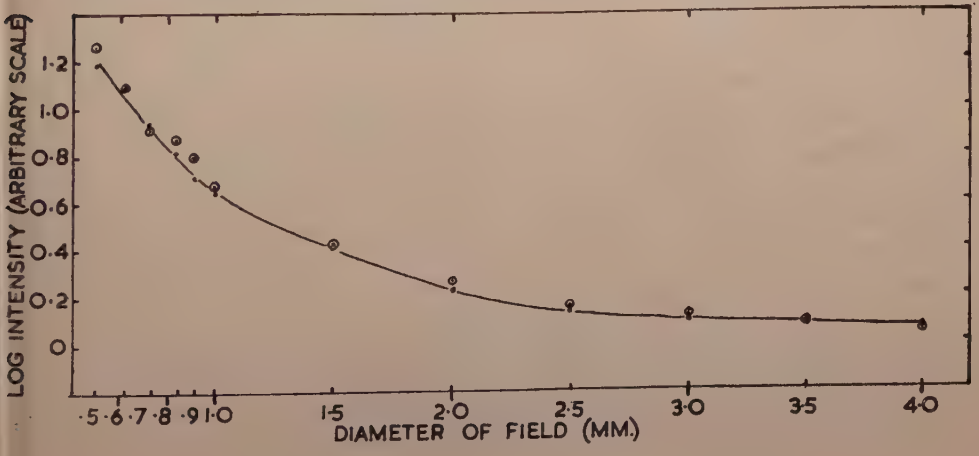
III

Results

The first experiments were designed to test whether violet or red test fields made any difference to the relationship between intensity and field-size. The threshold intensities were recorded with central fixation for each size of field from 2·4′ to 19·4′, and in order to eliminate day-to-day fluctuations in absolute threshold the logarithms

FIGURE 1

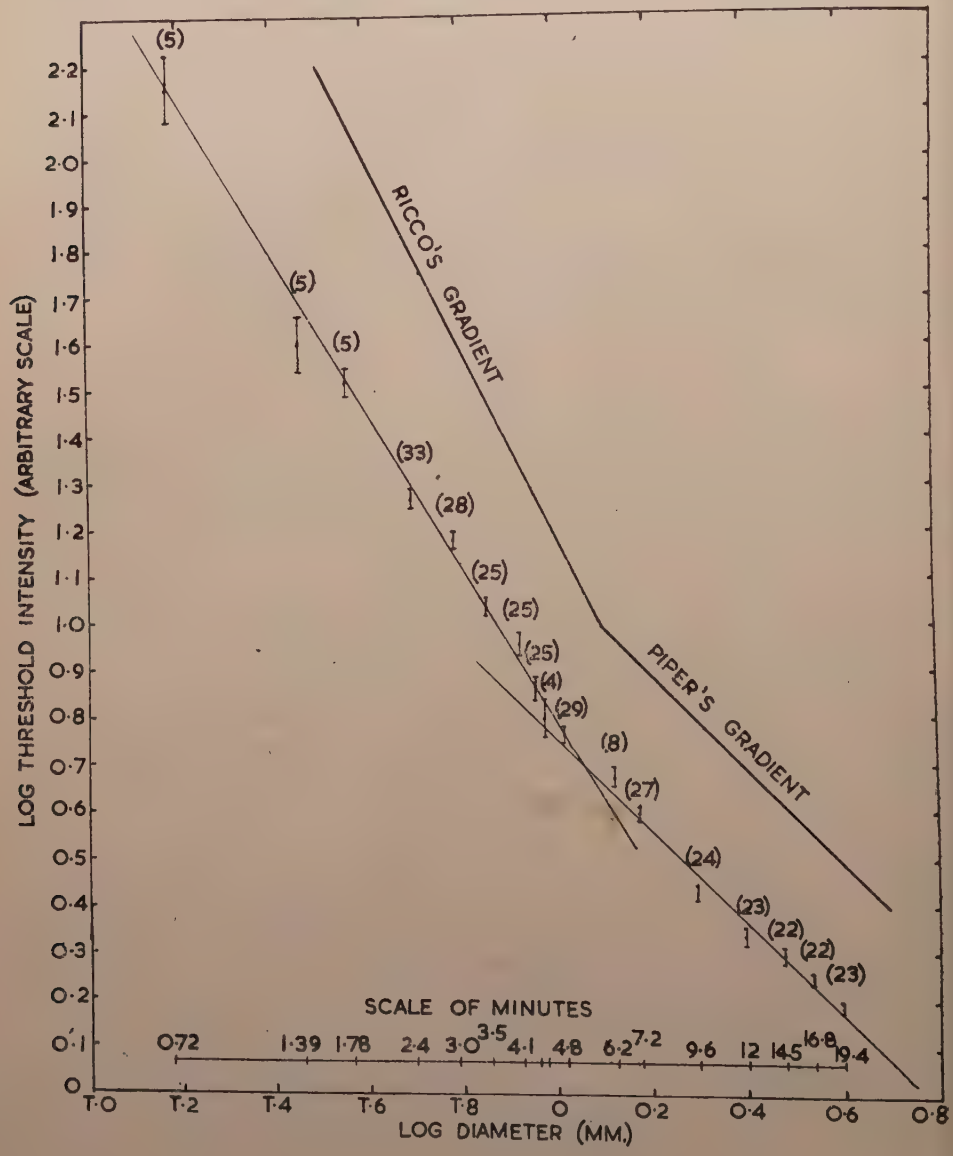
The Logarithm of the Intensity of the Threshold Stimulus in Arbitrary Units, Plotted Against the Diameter of the Field in mm., with Violet Light (⊙) and with Red Light (•).



of the intensities were plotted against the diameter of the field and the readings for each day were moved along the vertical axis in such a way as to obtain the best fits with preceding readings. The results are recorded in Figure 1. The intensity values are in arbitrary units. From this figure it is clear that very much the same relationship holds good for both the violet and the red test fields. The range of the

FIGURE 2

THE LOGARITHM OF THE THRESHOLD INTENSITY PLOTTED AGAINST THE LOGARITHM OF THE DIAMETER OF THE TEST FIELD IN MM. THE MEAN OF THREE SERIES OF READINGS ON TWO OBSERVERS; THE NUMBERS IN BRACKETS INDICATE THE NUMBER OF READINGS. THE VERTICAL LINES INDICATE THE STANDARD DEVIATION OF THE MEAN.



individual points upon which the points plotted in Figure 1 are based is of the order of 0·2 log units, so that the position of the circles (violet light) is nowhere significantly different from that of the points (red light). If Baumgardt is right in his hypothesis with regard to the interpretation of Ricco's and Piper's laws then it is clear that whatever relationships are established for the "red" receptors are equally valid for the "violet" receptors so long as vision is confined to the central fovea; it is therefore impossible to separate the two types of receptor by their behaviour with regard to threshold intensity and area stimulated, at least with field sizes down to 2·4′ in diameter.

TABLE I

Log diameter of field	*Angle subtended by field in minutes*	*Log threshold intensity, mean values (numbers of observations in brackets)*	*Standard deviation of the mean*
$\bar{1}.18$	0·72	2·15 (5)	0·07
$\bar{1}.46$	1·39	1·60 (5)	0·06
$\bar{1}.57$	1·78	1·52 (5)	0·03
$\bar{1}.70$	2·4	1·27 (33)	0·02
$\bar{1}.79$	3·0	1·18 (28)	0·019
$\bar{1}.86$	3·5	1·04 (25)	0·016
$\bar{1}.93$	4·1	0·96 (25)	0·024
$\bar{1}.96$	4·4	0·87 (25)	0·020
$\bar{1}.98$	4·6	0·80 (4)	0·04
0.02	4·8	0·77 (29)	0·014
0.13	6·2	0·68 (8)	0·020
0.18	7·2	0·60 (27)	0·016
0.30	9·6	0·43 (24)	0·014
0.40	12·0	0·34 (23)	0·022
0.48	14·5	0·30 (22)	0·016
0.54	16·8	0·25 (22)	0·013
0.60	19·4	0·19 (23)	0·010

The further question, however, remained whether the results are more adequately described in terms of some continuous relationship as suggested by Wald (1938) and by Graham, Brown and Mote (1939), or whether there is in fact a change from something akin to Ricco's law when the fields are very small to something more resembling Piper's law when the fields exceed a certain critical size, this size then representing the size of the quasi-independent unit as suggested by Baumgardt. Figure 2 represents the mean of three sets of results relating the logarithm of the intensity of the stimulus to the logarithm of the diameter of the field in mm. The figures (see Table I) were obtained from two normal observers, one set for violet light and the other two sets for red light. There is, in spite of the fact that most of the points are based on the means of more than 20 readings, some considerable margin of error in the exact location of the points and it would be possible to draw a continuous curve through the points as plotted. Nevertheless, there is a rather strong suggestion that the two straight lines drawn in the diagram might fit the points almost equally well. If this suggestion is accepted it would mean that one relationship holds for fields smaller than about 5′ or 6′ (log. 0·06 mm.) and another for fields larger than this. The other two lines on Figure 2 define the gradients

corresponding to Ricco's and Piper's laws, and clearly show that, between 5′ and 20′, Piper's law is very nearly true. With diameters less than 5′, however, there is a very considerable divergence from Piper's law and an approach to Ricco's law, though the agreement is not so satisfactory.

The position therefore seems to be that, while the data may be capable of description by some continuous function, in effect, a rather pronounced inflexion occurs at a field size which subtends about 5′ or 6′ at the eye. Above this size, the relationship $\sqrt{A}\, I = C$ is substantially true, while below this size the relationship approximates more nearly to $A\, I = C$.

IV

Discussion

The results of this analysis thus, in some respects, confirm and extend those of Baumgardt (1947, 1948) and if he is right in supposing that the change from Ricco's law to Piper's law occurs as soon as more than one independent or quasi-independent unit is involved, then it would appear that the area occupied by such a unit in the central fovea corresponds to a field subtending about 5′ or 6′ at the eye. This is manifestly too large to correspond to a single receptor for which the figure would probably be more of the order of 30″. If therefore it has any anatomical counterpart, this is more likely to be the flat-topped bipolar cell or the ganglion cell. The dendritic tree of the flat-topped bipolar cell is of the right order of magnitude and corresponds to a field size of about 6′ though it is probably somewhat smaller than this in the foveal region. If the size of this anatomical unit were shown to be connected with the change of law, as these results perhaps suggest, it is pertinent to ask whether any similar change in the area-intensity relationship occurs with still smaller fields corresponding in size to the midget bipolar cells. Is the divergence from Ricco's law with the smaller fields seen in Figure 2 perhaps explicable on the grounds that both midget bipolars and flat bipolars are involved, and would the law become rigidly true at a field size corresponding to the midget bipolar cell? The present methods and results do not give the information necessary to answer such questions. In the first place, very few figures have been obtained for the three smallest field sizes and these with red light only, on account of the practical difficulty of obtaining a sufficiently intense source of violet light. Secondly, it is clear that very small errors in the size of the stops delimiting the fields, when these are of the order of 0·5 mm. in diameter have a very significant influence on the threshold values obtained for these small fields. Moreover, and this is probably the most important point of all, the optical perfection of the image on the retina becomes questionable at these small dimensions and the actual form of the image on the retina is a matter of theory and guesswork. It is clear therefore that these suggestions as to the correspondence between the figures for the smallest fields and the anatomical units concerned are little more than ideas for future research. The only justifiable conclusion from the present data is that with fields smaller than about 5′ or 6′ in diameter, the product of area times intensity is much more nearly constant than that involving the square root of the area. Whether the relationship becomes rigidly true for areas subtending 1′ or so, cannot be determined from the data available. The fact that the square root relationship is so nearly true for fields subtending between 5′ and 20′ and then diverges so widely with smaller fields does perhaps indicate a critical point corresponding to fields of about 5′, and this would support the idea that some receptive unit in the central fovea has approximately these dimensions. It is clear, however, that the figures here presented for

the area-density relationship in the central fovea allow no decision to be reached as to whether in fact this apparent break at about 5′ is anything more than an inflection in a continuous curve relationship.

The author is indebted to the Leverhulme trustees for a grant enabling this investigation to be completed. He also wishes to express his gratitude to the two subjects (D. G. and P.M.H.R.) for submitting to the ordeal, and to W. A. H. Rushton for much constructive criticism.

V

References

Baumgardt, E. L. M. (1947). *J. gen. Physiol.*, **31**, 268.
Baumgardt, E. L. M. (1948). *C.R. Soc. Biol., Paris*, **142**, 464.
Graham, C. H. and Bartlett, N. R. (1939). *J. exp. Psych.*, **24**, 574.
Graham, C. H., Brown, R. H. and Mote, F. A. (1939). *J. exp. Psych.*, **24**, 555.
Konig, A. (1894). *S.B. Akad. Wiss., Berlin*, 577.
Polyak, S. L. (1941). The Retina. Chicago.
Wald, G. (1938). *J. gen. Physiol.*, **21**, 269.
Willmer, E. N. (1944). *Nature, London*, **153**, 774.
Willmer, E. N. (1949). *J. Physiol.*, **110**, 377.

(*Received 12th December*, 1949.)

AN EXPERIMENTAL STUDY ON THE FITNESS OF SIGNS TO WORDS

BY

K. R. L. HALL and R. C. OLDFIELD

(*From the Institute of Experimental Psychology, Oxford*)

This experiment, following Bartlett's Picture-writing, was designed to show: how much agreement there would be in the choice of a sign as the most fitting in connection with a word; whether the reasons given for a choice of sign would vary greatly, and would be rationalizations rather than a primary ground for so choosing; and whether the chosen signs were more easily recalled than the not-chosen.

Six cards were shown to the subject, on each of which was one word and two signs. The signs were not intended to represent any object, although none was arbitrary. The subject had to choose the more appropriate sign and give the reason for his choice. Afterwards he had to reproduce all the signs.

Agreement in choosing one of the signs of each pair was surprisingly high, although the reasons varied considerably. The chosen signs were reproduced more accurately and frequently, the reproductions showing the conventionalizing effect of the reasons given.

Choice of a sign as most fitting seemed to depend upon certain formal structural qualities of a general character. The most fitting sign for a verbal setting seemed to combine a conventional structure with simplicity and regularity of design. Linking these results with those of an experiment of Harrower's, it is considered possible that a method may here be available for exploring forms of mental organization not so far tapped by standard intelligence tests.

I

Introduction

In his Picture Writing experiment, Bartlett (1932) used a variety of signs which were more or less arbitrarily selected or constructed, and the subjects were required to use these signs in place of the words with which they were associated. Some of the signs were directly representative, some were designed to awaken secondary associations and some were "of the type commonly, but inaccurately, called 'meaningless'" (p. 97). As he pointed out, however, obviously no single classification could be made owing to the fact that every subject would classify such material in his own way. The general task was to learn the signs in connection with the words accompanying them, and, when the words were read out later on, to reproduce the signs they had learned for them.

In the "effort after meaning" required by the task and the nature of the material, a variety of reactions were observed in the subjects. The representative signs were "easy," and the connection obvious, but where a subject was unable to give a sign any precise and unambiguous representative significance, it would almost certainly be omitted in the reproductions.

Where a common and conventional setting was not provided by the word with which the sign was connected, an effort was frequently made to find some alternate form of setting. This might be achieved by finding some familiar secondary association, or some particular connection which could not be easily formulated in words but which was felt to be appropriate "in that peculiar and unmistakable way which always occurs when some presented material seems to be singularly in

its right place but we do not know why it should be so" (p. 113). For example, for the sign associated with the word "thing," one subject appeared to "feel" its particular suitability and pleasingness, and the sign for "Philosophy" was reported by several subjects to be peculiarly fitting.

Elaborating on this affective fittingness, Bartlett observed (p. 114) that the satisfaction which accompanied such an experience could occur both when the subject was able to say why the sign fitted, and when he was unable to formulate any reason. Two of his subjects had, for instance, found this same "Philosophy" sign particularly pleasing "simply because it was a big structure tottering on a little foundation." He concluded that such fitness of certain material may involve complex cognitive processes which do not need to be formulated in order to be appreciated.

Now these observations of Bartlett's were more or less incidental to the main purpose of his experiment, and it seemed that there might very well be important individual differences identifiable here in the reactions to similar sorts of material. The less "familiar" and "conventional" the material, the more probable it would be that such subjective differences might be observed.

We can therefore summarize certain of the implications of these findings as follows:

(1) The fitness of a sign in connection with a word need not be on account of any *representational* appropriateness or secondary associative significance.
(2) The peculiar fitness of a sign may be manifested by an affective reaction on the part of the subject. Any associative significance that may afterwards be formulated may therefore not necessarily have been operative in this primary "feeling" reaction. The formulation may therefore be in the nature of a rationalization.
(3) It is possible that individual differences in the way in which subjects deal with such sign-word combinations may be important where the connections are not conventional and easy.

The present experiment was therefore designed to show:

(1) How much agreement there would be, in a group of subjects of similar background, in the choice of a sign as the most fitting in connection with a word.
(2) Whether the reasons given for choosing a particular sign would show much diversity, and would seem to reflect more a subsequent rationalization than a primary ground for the choice.
(3) Whether the chosen signs were more easily and accurately remembered than the not-chosen.

II

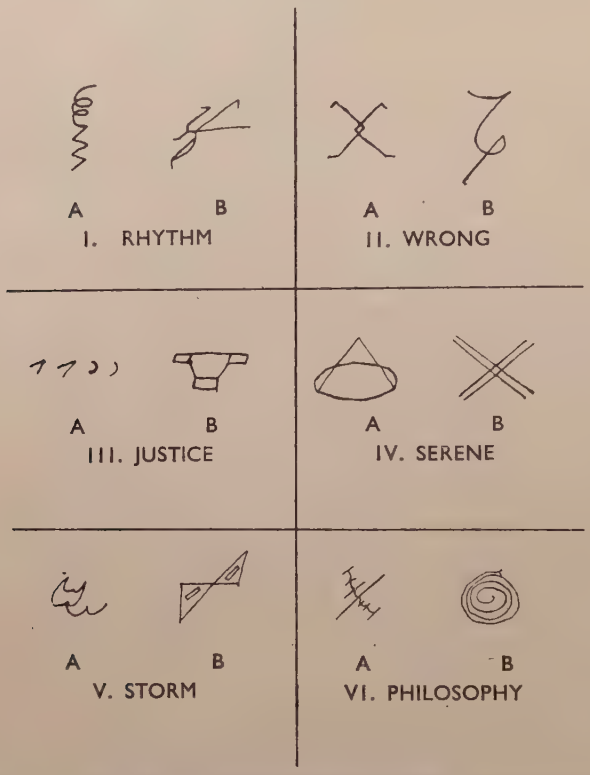
Description of the Experiment

The Method of Choice was used, that is, a pair of symbols was presented in connection with a word, and the subject's task was to choose which of the two he thought to be the more appropriate.

The materials used consisted of six plain white cards, 3 by 5 in., upon which were drawn in black ink two symbols or signs, under which was printed a word. The set of twelve signs and the six words which accompanied them are shown below. Two of the signs were the same as those used by Bartlett in his Picture-writing experiment, sign A for Philosophy, and sign A for Wrong. In drawing the remaining ten signs, it was intended that they should be similar to those that Bartlett had used in his third set, and direct object-association was avoided as far as possible. They were, however, drawn

in connection with the words they accompany, that is, they were not arbitrary, and not randomly selected, but there was no intention of making one or other of a pair the more significant member.

FIGURE 1



The subjects who took part in the experiment were all of university educational standard, both male and female, and represented a fairly wide range of intellectual background.

There were three main parts to this experiment, in each of which the experimental conditions were slightly varied. The essential task was the same, that of choosing a sign with reference to a word, throughout the experiment.

Preliminary. 17 subjects—8 Male, 9 Female.

This was mainly to help in deciding the best experimental procedure. Each of the six cards was presented to the subject for 10 seconds, the order in which they were presented being altered from subject to subject. The subject was instructed to choose the A or the B symbol, according to which he thought was the most fitting or meaningful, but might say "neither" or "both," if he was unable to judge between them. He was asked to state the reason for his choice.

Fourteen of the 17 subjects were asked, immediately after they had finished their choices, to reproduce both the symbols which they had been shown in connection with a particular word. The six words were printed on a sheet of paper, and the reproductions were made in an A and B column opposite. There was no time limit to this part of the experiment, and no previous indication had been given to the subjects that they would be required to recall what they had seen on the six cards.

Main Experiment. 40 subjects—29 Male, 11 Female.

(1) Using the same material, the procedure was modified. The subjects were instructed definitely to choose either A or B, and no time limit on the making of the choice was imposed. Where a subject could not decide between A and B,

he was eventually asked to guess, as this was found invariably to elicit a choice, usually backed by a reason.

(2) One variation was introduced, in order to determine whether the position of the signs on the card might be a factor in biasing the choice, and, for 10 of the subjects, a second set of cards was used, with the B sign of the original set on the left and marked A, so that the position was reversed. The procedure was identical.

(3) Ten subjects were asked for reproduction of the symbols, as in the Preliminary Group, except that here there was an interval of 10 minutes between making the choices and the reproductions. Seven other subjects were also asked to reproduce the symbols, the delay being varied from five minutes up to three weeks. For two other subjects, the reproduction procedure was reversed, and, instead of being given the words, the 12 signs were shown singly, in random order, and the subject was required to respond with the correct word for each.

Group III. 20 subjects—18 Male, 2 Female.

In order to find out whether certain of the symbols were predominantly chosen irrespective of the verbal context in which they were framed, that is, whether, in at least some cases, the relationship between word and symbol was much less specific than might have been supposed, the same pairs of symbols were shown, but with the words blanked out so that they were not in any way visible. The subject was instructed as follows: "There are six cards here, each with two symbols drawn on them. Before I show them to you, I am going to tell you a word, and, when I show you the first card, I want you to consider the two symbols on it in relation to that word, and choose which of them is most fitting and meaningful in that connection. With the same word in mind, I want you to choose one of the two symbols on each of the other five cards."

When all the six cards had been shown, with reference to the first word, they were spread out on the table in front of the subject, and he was asked to say which symbol he thought was the best in connection with that word.

The same procedure was carried out until all the six words had been used, and choices from every pair of symbols made in connection with each. The order of the words was varied from subject to subject, as was the order of presenting the symbols in connection with the words.

The subject was not asked to give any reason for his choice.

III

Results

Distribution of Choices

(1) *General Distribution.*

For each of the six cards, a very substantial measure of agreement of choice was found. Table I shows the actual distribution of choices made, for the 40 subjects of the main experiment.

TABLE I

Card	*Number choosing A*	*Number choosing B*	*P*
1	34	6	$\ll 0.01$
2	31	9	$\ll 0.01$
3	8	32	$\ll 0.01$
4	31	9	$\ll 0.01$
5	33	7	$\ll 0.01$
6	3	37	$\ll 0.01$

We may examine this distribution relative to the hypothesis that neither figure on each card is more likely to be chosen than the other. Column P in Table I shows the probabilities of the observed distributions, or their opposites (viz. (40–A), (40–B),

in place of A, B), occurring on this hypothesis. It is clear that the agreement actually found is far in excess of that to be expected if only chance factors were involved. It does not follow, of course, that the agreement is due wholly to the fittingness or otherwise of the words.

(2) *Figure Position.*

Since it seemed possible that the choice might be biased by the position in which it was drawn on the card, or the letter by which it was designated, 10 subjects were shown cards in which the position of the figures was reversed, but the left-hand one was still labelled A and the right B. Table II shows the choice distributions for these 10 subjects (Group II) as compared with the choices made by the remaining 30 (Group I) subjects. The columns A and B give the choices for the actual figures as shown in Figure 1, irrespective of whether they were labelled A or B. Values of chi square are shown for the table—Group I: Group II, A : B, and it is clear that in no case is there any evidence of figure position or letter-label affecting choice-distribution.

TABLE II

Card	*Group*	*A*	*B*	*Chi square*
1	I	26	4	0·26
	II	8	2	
2	I	23	7	0·05
	II	8	2	
3	I	6	24	0·00
	II	2	8	
4	I	25	5	2·38
	II	6	4	
5	I	24	6	0·51
	II	9	1	
6	I	2	28	0·36
	II	1	9	

In view of this, Groups I and II have been pooled in the further consideration of the results.

(3) *Sex Difference.*

Our 40 subjects included 29 men and 11 women. Table III shows the choice distributions according to sexes, and it will be seen that there is no significant difference between them.

TABLE III

Card	*Sex*	*A*	*B*	*Chi square*
1	M	26	3	1·79
	F	8	3	
2	M	23	6	0·20
	F	8	3	
3	M	6	23	0·11
	F	2	9	
4	M	22	7	0·20
	F	9	2	
5	M	25	4	1·00
	F	8	3	
6	M	3	26	0·05
	F	0	11	

(4) *Effect of Instructions* (Preliminary Group).

As was seen in connection with the preliminary part of this experiment, somewhat different instructions were given. If the subject was unable to make a clear choice he was not asked to guess, but was allowed to say that neither figure stood for the concept in question, or that both did. In these conditions, the choice distribution shown in Table IV was found.

TABLE IV

Card	*A*	*B*	*Both*	*Neither*
1	9	7	1	0
2	13	1	1	2
3	2	13	0	2
4	7	7	0	3
5	10	3	2	2
6	2	11	2	2

With these instructions, then, 17 per cent. of the responses are either "both" or "neither," and a clear choice is only obtained in 83 per cent. But, in addition, the different instructions appear, at least in the case of cards 1 and 4, to have produced a markedly different choice distribution even where a clear choice was obtained. Table 5 shows the values of chi square found for the choice-distributions in the two cases, "both" and "neither" responses being neglected.

TABLE V

Card	*Chi square*	*Card*	*Chi square*
1	5·30	4	3·77
2	1·62	5	0·20
3	0·33	6	0·71

The value of chi square for card 1 is significant (using P = 0·05 as criterion), while that for card 4 might be expected less than once in 19 times by chance, though not less than once in 20, and so is only just below the conventional level of significance. It seems, therefore, that the *general* attitude of the subject may exercise an influence upon the *particular* choice made in respect of the character of fittingness. This clearly suggests that caution should be exercised in drawing conclusion from experiments in this field by methods involving choice. The numbers in this group are, of course, small, as compared with the 40 subjects of the main experiment.

(5) *Distribution of "Unpopular" Responses.*

TABLE VI

Number of unpopular responses	*Number of subjects*	*Number of subjects*	*Number of subjects*	*Total number of unpopular responses made*
	Male	*Female*	*Total*	
0	14	3	17	0
1	7	3	10	10
2	4	5	9	18
3	3	0	3	9
4	0	0	0	0
5	1	0	1	5
6	0	0	0	0
	29	11	40	42

Inasmuch as, in the result for Group I and II, there is a clear and significant consensus of choice, we may speak of the figure chosen less frequently as the "unpopular" figure, and consider the distribution of unpopular responses among the subjects. Table VI shows this distribution.

Among the male subjects, in fact, there were 14 out of 29, or nearly 50 per cent., who gave no unpopular responses at all.

Reasons for Choice.

All the subjects in this experiment were asked, when each choice had been made, to give the reason for their decision. In a few cases (22 out of 240 for Groups I and II) they were quite unable to do so. In the remainder they were able to give *some* reason, although often with a good deal of hesitation, so that one gained the impression that the choice itself was a direct schematically determined reaction like the act of recognition, while the evocation of the reason was less a statement of the causes for the choice than a *justification* of an attitude already evoked in the choice.

One figure might be chosen rather than another, because:

(1) of some alleged similarity of quality or form between figure and idea (such as circularity for philosophy),

(2) it suggested some form or object connected with the idea and, in this case, the object or form might be either—

  (i) one conventionally regarded as a symbol for the idea in question (scales or balance for justice), *or*

  (ii) one in some other way connected with the idea (judge's wig for justice, cloud for storm),

and here the association is similar to the conventional object except that it is *actually* to be observed in this connection, whereas the scales of justice are only so shown in pictures or on statues, and so on.

A very small number (7 in 240) of cases were found in which a figure was chosen because it suggested some *activity* connected with the idea (e.g. dancing, for rhythm). And, apart from those cases in which no reason could be given, a few (20 in 240) were unclassifiable under any of the above heads.

Summarizing some of the main points from the reasons given for the choices, it was found that:

(1) It is some quality, whether of the figure as it stands, unnamed, or whether, of a named, associated object, such as scales, that is the commonest apparent factor in determining the choice. It seems probable, for instance, that it is the *balance* of the popular figure for Justice, rather than any realistic resemblance to a pair of scales, which is important for the choice; that, in fact, it may be a general quality of the figure rather than any specifiable associated object that is influential, for, as is seen in more than one instance, object-associations can perfectly easily be found for the unpopular choices, and it is not, therefore, a matter of objective resemblance.

(2) The quality of the figure, in terms of symmetry, balance, regularity, and so on, may, where the word accompanying may suggest it, actually serve as a ground for *not* choosing that figure.

(3) In a few instances, subjects tend to avoid the obvious association, as in the cross-out sign for Wrong.

(4) There is, particularly for card 5A, a preference for the less stable, and more "moving" figure in connection with storm: this, again, is a perception of a general quality that seems to count most rather than perception of any particular object-likeness—for the very variety of objects evoked is sufficient evidence of the generality of the figure.

(5) There is no very overt indication of affective reaction, such as Bartlett found, yet the general impression derived from observing the subjects' reactions and from some of the verbalizations regarding the quality and symmetry of a particular figure is that the fittingness of, say, the B figure for Philosophy or the A for Serene is much more *felt* than logically conceived.

Conclusions on the Main Experiment.

(1) In any particular choice set up there may be a very high degree of agreement between different people of the same kind of group about the relative fitness of one or the other figure.

(2) The grounds given for the choice may be very various in their wording, even where there is considerable agreement on choice, and even in the type of reason given, although there may be a common and more general ground for choosing a particular figure discernible through this variety.

Group III.

As a final variation in the choice procedure, arising out of the results just considered, a further group of 20 subjects was shown the six cards with the words blanked out, the procedure being as described on page 63. It was thought that one sign might be chosen rather than another, because it is, in itself and irrespective of its verbal context, especially pleasing, and also that one sign might be considered particularly suitable in connection with several words. For reasons of space the detailed results are not shown, but, from the results it seems that the general impression of the main experiment is confirmed, namely, that it is some formal, qualitative aspect of the sign that makes it either chosen or rejected as the case may be, rather than any object-association being the *primary* ground for choice. A sign is not absolutely fitting in connection with any verbal setting, but is selected in its comparative relation to the other member of the pair, making up the total problem-unit.

Reproduction of the Figures.

Fourteen subjects of the Preliminary Group made reproductions of the figures immediately after they had completed the choice part of the experiment, and 10 subjects of the Main Experimental Group made reproductions after an interval of 10 minutes. As has already been emphasized, the experimental task was specifically stated to be the choosing of the best figure, and no expectation of having to recall the figures was aroused.

TABLE VII

Subject group	*Number of subjects*	*Chosen*		*Non-chosen*		*Value of chi square*	*P*
		R.	*N.R.*	*R.*	*N.R.*		
Preliminary	14	79	5	42	42	40·43	$\ll 0{\cdot}01$
Main	10	56	4	20	40	46·50	$\ll 0{\cdot}01$

Table VII gives the number of chosen figures that were reproduced and the number of non-chosen figures reproduced for these two groups. Although there was considerable variation from the originals in the reproductions made, it was in almost

every case clear which figure was intended. Only the *correct* reproductions (R) are counted in Table VII, and those which are clearly incorrect and which show no recognizable similarity to the originals are classed with the not-reproduced (NR).

The difference between the number of chosen and the number of non-chosen figures reproduced is highly significant for both groups of subjects, while there is no significant difference obtained from comparing the Preliminary Group frequencies with the Main Group, the value of chi square being 0·0024, which, for one degree of freedom, gives a probability far higher than 0·05. Thus, approximately twice as many of the figures chosen by the subjects as the more appropriate are reproduced with reasonable accuracy.

Regarding the actual number of reproductions made for each figure, these are shown in relation to the number of choices made for that figure in Table VIII. For the Preliminary Group, where the judgment was "neither" or "both" in choosing for a card, this has been omitted from the number of choices shown, only definite A and B choices being counted. Hence the totals do not come to 14.

TABLE VIII

Preliminary Group

	I		*II*		*III*		*IV*		*V*		*VI*	
	A	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
Chosen	8	5	11	1	2	10	6	5	8	3	2	9
Reproduced	10	8	11	6	7	13	10	12	13	10	5	14

Main Group

Chosen	9	1	9	1	2	8	8	2	7	3	0	10
Reproduced	10	4	10	3	5	9	7	7	7	3	2	10

There is clearly a close relationship between Choice and Reproduction, particularly with the Main Group, where judgment was required to be a definite A or B. For this Group, only for Card IV is the number of reproductions equal for each figure, and this is probably due to the fact that some subjects gave a negative reason for choosing A, by saying that B is cross, or a busy cross-roads, and therefore is the opposite of Serene. The verbalization which the subject made very often influenced the reproduction of the figure to which it referred. It will be observed that Figures I.A, II.A and VI.B are particularly privileged both in Choice and Reproduction for this Group.

Qualitative Changes in the Reproductions

(1) *Conventionalization.*

Many of the changes in the reproductions, when compared with the original figures, were clearly due to Object Assimilation or Conventionalization. The subject was required to give his reason for making a choice, although no attempt was made to influence him if he was unable to formulate the reason, and some subjects rested content with giving no particular reason or only a very vague statement of it.

Considered in conjunction with these verbal reasons, the effect of the verbalization upon the reproductions is obvious, and, in some cases, quite striking. Figures such as I.A for Rhythm, II.A for Wrong, and VI.B for Philosophy, apparently lend

themselves very readily for easy assimilation and reasonably accurate reproduction. Certain small details are omitted, and the II.A figure might become simplified to a conventional cross-out sign, I.A might be perceived and recalled as a "regular, smooth-flowing spiral," or as a coiled spring, so that the jagged, lower part of the figure is reproduced as circular, in conformity with the upper part.

A greater variety of assimilative changes can be seen in the reproductions of III.A and V.A, in keeping with the variety of object association that these figures aroused. Hence the reproductions are on the whole less accurate than those for I.A or VI.B, for example. If the same figure is chosen either because it is like a judge's wig, a gallows, a court bench, or scales of justice, the reproductions may be expected to show corresponding deviations.

There is, therefore, no doubt whatever that the verbalization intervening as a reaction between the making of the choice and the making of the reproduction has in many cases influenced this reproduction. The name given, whether it be an object-name such as "a man dancing" or "an empty cross-roads," or a more general name such as "spiral," "circularity," or "zig-zag," is often that which is aroused when reproduction is required. The conventionalizing effect of the name, which Bartlett, Wulf and Gibson observed, is confirmed, while there is further evidence of the simple and symmetrical characteristics of these conventional, conceptual patterns to which the figures are related.

It seems, therefore, that the factors which influence the original judgment of fitness in the choice of a figure in connection with a verbal setting serve to enhance the recall-value of that figure. Simplicity and economy of structure, within the framework of the verbal setting, are both preferred in perception and achieved in reproduction.

(2) *Figure Assimilation.*

This type of change that both Bartlett and Gibson observed was also found in a few cases. So, in three reproductions made of Figure I.B, there is evidence of confusion with Figure II.A.

IV

Discussion

Where, as in the material used in the present experiment, one is dealing with more or less complex relationships between words and symbols, the nature of the real basis of choice does not reveal itself very clearly. The high number of choices falling on one particular sign rather than the other is not supported by a corresponding agreement in the reason for making that choice. What then is the nature of this fittingness of word and sign?

Apart from Bartlett's own experiments, little work seems to have been done on this problem. Harrower (1933), however, carried out an interesting experiment when studying "Organization in the Higher Mental Processes." She thought that "the joke allowed of some sort of diagrammatic representation of the particular relations holding within its parts: directly analogous to the difference in form and structure in the fields of perception" (p. 58). She accordingly presented a set of four diagrams with each joke, three of which were arbitrary, that is they were not intended to bear any relation to the joke, while the fourth was intended to express the relational *structure* of the joke. She found that, not only were most of her subjects able to select the correct diagram of the four, but that the recall value of the jokes when accompanied by the proper diagram was far higher than where the same jokes were accompanied by arbitrary diagrams, and far higher, also, than where

the joke was presented unaccompanied by any diagram at all. The explanation of this was that the structure of the joke was enhanced by the presence of the diagram, and this structure kept the parts of the joke well articulated, just as any clearly apprehended relational scheme might do with other varieties of material.

Our own results, in the reproductions of the signs, seem to confirm the superiority of recall-value of the better articulated name-sign combination. In either case, it seems that there has to be an active effort on the part of the subject to make the material meaningful. Taking these findings of Harrower's into consideration, there seems little doubt that the grounds of fittingness for certain of the signs used in the present experiment may be described in terms of structural relevance rather than object-association. The words Philosophy, Wrong, Justice, etc., are general, and the basic concepts for which they stand may be supposed to be devoid of all object-detail and particularity. It seems, further, as though the most appropriate sign must not only have a certain relational significance, but also a formal economy and simplicity of the Gestalt kind.

The question of the fittingness of such signs or diagrams to verbal settings appears to raise many problems which go beyond both the experimental and theoretical scope of the present enquiry and into the psychology of thinking. It seems possible, for instance, that this method may offer a means of evaluating the qualitative structure in an individual's verbal intelligence, so that what he has made of his experience can be objectively assessed.

In conclusion, the present results seem to bring out more clearly the nature of the affective fittingness that Bartlett had observed. The choice of a sign as the most fitting in relation to some general verbal setting appears to depend, not so much upon the possibility of associating it with some object, as upon certain formal qualities of simplicity and regularity of outline combined with a schematic, structural relevance to the particular concept. That the "*feeling* of fittingness" may be primary in the process of choosing, there seems little doubt, and the reason given is often secondary—a rationalization of a more immediate perceptual process.

V

Summary and Conclusions

(1) The choice of a sign as fitting a particular verbal setting appears to depend, not only upon the relationship it bears to the other sign or signs with which it is compared, but upon certain formal, structural qualities of a general character.

(2) It seems probable that it is not, therefore, any object-association, such as is apparent in the verbalizations, which is the primary ground for choice, but rather this structural correspondence between figure and verbal setting.

(3) Within the limits of the material used, it seems justifiable to conclude provisionally that the most fitting sign for a particular verbal setting will combine a conventional structure with a general simplicity of design.

(4) It is, perhaps, in bringing out these structural aspects in the perception of signs in relation to words that the present results have made some contribution to the problem that Bartlett's experiment suggested. Where words are involved in the process of Conventionalization, the generalization apparently necessary may involve both simplification and regularization of form.

VI

References

Bartlett, F. C. (1932). *Remembering.* Cambridge.

Gibson, J. J. (1929). The reproduction of visually perceived forms. *J. exp. Psychol.*, **12**, 1–39.

Harrower, M. R. (1933). Organization in higher mental processes. *Psychol. Forsch.*, **17**, 56–120.

(*Received* 31*st January*, 1950.)

STIMULUS INTENSITY AND ATTENTION IN RELATION TO LEARNING THEORY

BY

D. E. BERLYNE

(*University of St. Andrews*)

Three experiments are described which test the hypothesis that the more intense of two stimuli will, *ceteris paribus*, be more likely to receive attention. It is assumed that an objective behavioural manifestation of attention to a given stimulus is a preference for responding to it rather than to another which is present at the same time.

In all three experiments, successions of pairs of visual stimuli interspersed with single stimuli were presented to the subject, and he was instructed to respond to either (by pressing its corresponding morse-key), but not both, in the case of the pairs. The first two experiments reveal significant tendencies to respond to the larger and the brighter stimulus respectively. In the third experiment, there was a tendency, but a statistically insignificant one, to respond to a constant rather than to a flickering stimulus.

It is shown that the attraction of attention by a more intense stimulus follows from Hull's system with the addition of his new variable, "stimulus-intensity dynamism (*V*)," and it is suggested that it may thus be possible to add attention to the phenomena that can be integrated with an objective behaviour theory.

I

INTRODUCTION

The experiments to be reported in this paper* were part of a series designed to throw some light on the problem of interest. As the writer has pointed out in a previous publication (Berlyne, 1949), one aspect of behaviour to which the term "interest" has been applied is the preference of the organism for responding to a particular stimulus or group of stimuli when confronted with several. Bartlett, for instance, defines "an interest" as "a preferential response towards certain stimuli or objects or topics" (Bartlett, 1939). Moreover, interest has always been regarded as something closely bound up with attention, and one way, possibly the principal way, in which we can judge from overt behaviour which part of the perceptual field a subject is attending to is to note which part is prepotent in determining his responses.

This is, of course, precisely what happens in the familiar introspective experiments on attention. We present, perhaps tachistoscopically, a display which is too complex for the subject to apprehend the whole of it in the short time for which it is exposed. Then we ascertain what parts or aspects of the display he is capable of describing and conclude that those are what he has been attending to. Now, at one time or another he has learnt to name or describe elements similar to all those shown him; each of them has a particular verbal response which it could evoke. But only those to which he has "paid attention" have actually dominated the situation sufficiently to impose on him their own, even if, as the Gestalt school would remind us, the others might not have been without influence. They have, in other words, been prepotent. Similarly, in studies of fluctuating attention the stimulus in question alternately is and is not capable of calling forth its particular response and overcoming competing response-tendencies. But unfortunately, this interpretation in terms of objectively observable phenomena has usually not been used and most experimenters have worked with relatively complex visual material and with

* These experiments were performed at the Cambridge Psychological Laboratory with the help of a Rockefeller Research Grant.

the verbal response which entails so many complicating factors. We have as a result been prevented from linking attention with other aspects of behaviour and seeing the relevance to attention of the principles that recent learning theory has found to underly all forms of response.

To study attention without the uncertainties of the introspective method, it is therefore desirable to devise as simple a situation as possible in which the subject is presented with more than one stimulus at once but is required to respond to one only. Many well-known studies involving the use of two or more simultaneous stimuli have been made, but they appear to fall into one or other of the following categories:

(i) Which stimulus is prepotent is made to depend on learning and association with reward or punishment (e.g. Yerkes' discrimination-box, Lashley's jumping-stand and Hunter's delayed reaction).

(ii) It depends mainly on the present motivational state of the subject (e.g. Jenkins and Warner's obstruction-box and Brown's conflict-experiments).

(iii) It depends on previous verbal instructions (e.g. experiments on distraction).

Now, such factors as learning, motivation and *Aufgabe* undoubtedly influence attention profoundly, but it has been held by many writers and confirmed by common experience that other, more intrinsic qualities of the stimulus, particularly intensity and novelty, also do so (see, e.g. Titchener, 1901). Little has so far been done to study or analyse the effects of these.

The experiments to be reported are concerned with the question of stimulus intensity, as represented by the size, brightness and steadiness of visual stimuli. A series of pairs of stimuli, with single stimuli intermingled with them, is presented to each subject, and, in the case of the pairs, he must respond to one only. The situation is such that the instructions, learning and motivation cannot favour one choice rather than another, and so intensity is free to show its effects.

II

Experiments

Apparatus.—The subject sat facing a vertical wooden board, approximately 18 in. wide by 12 in. high, in which there were four square holes at about eye-level, each 1½ in. per side and 1½ in. from its neighbour, in a horizontal line. Just behind each hole was a 3·8 volt torch-bulb, and in front of it a moveable cardboard screen could be interposed. Each screen had a paper-covered aperture and the size, shape and colour of the stimulus could be varied by varying the size and shape of the aperture and the colour of the paper. Intensity could also be reduced by inserting a resistance in the circuit.

Below each hole at the base of the board was a morse-key, whose end projected through the board to show the experimenter, by means of a simple mechanical device, which key was being pressed.

The subject sat just near enough to the board to reach the keys comfortably.

Subjects were naval ratings and students.

Experiment 1

Procedure.—The four stimuli used in this experiment were two white circles, approximately ½ in. in diameter, and two *larger* white circles, approximately ¾ in. in diameter. A different permutation was used with each subject to cancel out the effects of position.

The subject was first told that he would be given a few practice-trials. One light at a time was switched on, and he was to respond by pressing down its corresponding key as quickly as possible, releasing it as soon as the light went off.

Then the experiment proper began. Forty-eight stimulus-patterns were given, consisting of six presentations of each of the four stimuli alone and four of each of the

six combinations of two. The order was random, but the same for each subject. The subject was instructed to respond by pressing the appropriate key in the case of a single stimulus but to respond to one and one only when two appeared. "It doesn't matter which, whichever you feel like pressing first!" Only the first and second fingers of the right hand were to be used, and these had to be returned to a certain spot opposite the middle of the board after each trial.

Half the pairs were irrelevant to the purpose of the experiment, since they consisted of two stimuli of equal size. But these and the single stimuli were introduced to increase the number of possibilities and thus make it harder for the subject to prepare himself, making the choice more immediate and spontaneous.

Results.—The only trials whose results are relevant are those with a presentation of two heterogeneous stimuli (i.e. 16 trials per subject).

Mean responses to larger stimulus: 13·27
" " " smaller " : 2·63
Number of subjects: 6.
$\sigma^2 = 9{\cdot}33 \quad t = 3{\cdot}9 \quad \nu = 5 \quad p = 0{\cdot}012.$

Experiment 2

Procedure.—As in Experiment 1, except that the stimuli were two white circles, $\frac{1}{2}$ in. in diameter, and two *dimmer* white circles of the same size, intensity being reduced by the insertion of a 10-ohm resistance in the circuit.

Results.— Mean responses to brighter stimulus: 13·5
" " " dimmer " : 2·5
Number of subjects: 6.
$\sigma^2 = 8{\cdot}75 \quad t = 4{\cdot}15 \quad \nu = 5 \quad p = 0{\cdot}009.$

Experiment 3

Procedure.—As in Experiment 1, except that the stimuli were two white circles, $\frac{1}{2}$ in. in diameter, and two similar circles with *flickering* lights. The flickering was produced by moving the switches.

Results.— Mean responses to constant stimulus: 10·2
" " " flickering " : 5·8
Number of subjects: 10.
$\sigma^2 = 10{\cdot}96 \quad t = 1{\cdot}99 \quad \nu = 9 \quad p = 0{\cdot}078$ (not significant).

III

Miscellaneous Observations

The following facts emerged from all experiments:—

(*a*) Almost every choice was made, according to introspective reports, automatically, without any effort, thought or deliberation, although the stimuli were generally perceived simultaneously.

(*b*) Introspections were frequently belied by the results. Subjects denied tendencies to prefer responding to the more intense stimulus when such preferences were clearly demonstrated by the results, and some asserted tendencies which did not exist. This has serious implications for the use of the introspective method.

(*c*) There were no striking effects of position on the responses and, in any case, randomizing prevented any positional preferences from invalidating the results.

The mean numbers of responses to positions were as follows (from left to right from the subject's point of view):

Experiment 1 ..	6·3	5·3	7·7	4·7
" 2 ..	5·0	5·3	8·0	5·7
" 3 ..	4·2	6·6	6·9	6·3
Mean .. (22 subjects)	5·0	5·9	7·4	5·7

(All pairs, heterogeneous and homogeneous, are counted.)

(*d*) Some subjects showed the following tendencies:

(i) To adopt a systematic procedure, e.g. to respond always to the right-hand stimulus. This was more common with students than with sailors, and the results of such subjects were ignored, as this prevented any effects of the variable to be studied from showing themselves.

(ii) To respond to the stimulus just responded to or near it.

(iii) Rarely, to vary responses intentionally.

(iv) Rarely, to "compromise," by responding to a stimulus (not lit) between the two presented.

(v) To attempt, almost compulsively, to anticipate where the next stimuli will occur, although it was realized that they were randomized.

(vi) To feel anxious lest mistakes should be made, although mistakes (i.e. responses to stimuli not lit) were actually hardly ever made.

(vii) To develop stock responses for each combination. This would presumably have been more in evidence if the experiment had lasted much longer.

IV

Discussion

It was assumed that in these experiments learning and motivation would be kept constant so that the influence of the intensity factor would come out. The principal motivation, whether desire to conform to the experimenter's wishes, anxiety not to make a mistake, confirmatory reaction, etc., was satisfied equally whatever the choice of response. Since the response was made immediately after instructions, the only learning involved was learning to understand language, which can be assumed not to have varied for different stimuli. The results were also independent of any tendency to respond to position and can therefore be taken as due solely to intensity-differences.

Experiments 1 and 2 represent the familiar phenomenon of a larger or brighter stimulus "catching the eye" or "attracting the attention." The results confirm the hypothesis that a more intense stimulus is more likely to be attended to.

Experiment 3 was undertaken to test the common belief that a changing or moving stimulus likewise attracts attention. The results show a tendency, contradicting this expectation, but conforming to the above-mentioned hypothesis, to respond to the steadier stimulus, i.e., that producing more stimulation over a given period. This tendency was, however, not significant, which may be due to a conflict between two antagonistic factors, namely, the effect of stimulus intensity, favouring the steady stimulus, and another factor, favouring the flickering one. The latter is probably connected with the rôle of novelty and change in attention and may depend on a phenomenon which emerged from other experiments and is at present undergoing further investigation.

Now that the hypothesis may be regarded as confirmed, we must consider whether it can be related to fundamental principles underlying phenomena in other fields of psychology. It would appear that it can in fact be deduced from the system that, despite its undoubted inadequacies and provisional character, has so far succeeded in integrating the widest range of psychological phenomena on a rigid hypothetico-deductive basis, namely, Hull's behaviour theory.

Since these experiments were completed, Hull has introduced "stimulus-intensity dynamism (V)" as a new variable in his system (Hull, 1949). It now appears that our results, which were previously inexplicable by his learning theory, represent the effect of V with learning and motivation held constant.

Hitherto, "reaction-potential ($_sE_R$)," an intervening variable on which the latency ($_st_R$), probability (p), amplitude (A), and extinction-resistance (n) of responses depend, was treated as a function of habit-strength ($_sH_R$) and drive (D) (Hull, 1943).

But now it is stated to be dependent also on V, which is a variable increasing with stimulus intensity (i). The tendency to respond to the more intense of two stimuli can thus be shown to follow from the postulates of *Principles of Behavior*, with the addition of the new postulates concerning V:

(i) V is provisionally stated to increase with i (stimulus-intensity) according to the equation

$$V = A\ (1-10^{-b \log i})$$

where A and b are empirical constants.

The more intense stimulus will therefore have a higher V.

(ii) "Other things constant" (as they are in the conditions of our experiments), $V = {}_sE_R$ (reaction-potential). The response to the more intense stimulus will therefore have a higher ${}_sE_R$.

(iii) "Momentary effective reaction-potential (${}_s\dot{\bar{E}}_R$)" depends on ${}_sE_R$, $\dot{I}_R$ (inhibitory potential) and ${}_sO_R$ (an inhibitory factor which oscillates "according to the normal 'law' of chance") (Hull, 1943, Postulate 10). Since the last two factors, $\dot{I}_R$ and ${}_sO_R$, can be taken to be more or less equal for all response-tendencies in our experiments, the one with the higher ${}_sE_R$ will have the higher ${}_s\dot{\bar{E}}_R$ for most of the time.

The response to the more intense stimulus will therefore have a higher momentary effective reaction-potential (${}_s\dot{\bar{E}}_R$) more often.

(iv) "When the reaction-potentials (${}_sE_R$) to two or more incompatible reactions (R) occur in an organism at the same time, only the reaction whose momentary effective reaction-potential (${}_s\dot{\bar{E}}_R$) is greatest will be evoked" (Postulate 16).

The more intense stimulus will therefore be responded to more often. —Q.E.D.

The above postulates and their applicability can be checked by further experiments with different intensities. The probability of response to the more intense stimulus can be found for each pair of intensities, and it can be seen whether they are compatible with Hull's formulae. If they stand the test, the way may be open to the formation of something that has been long overdue, namely, an objective theory of attention.

Since the completion of this paper, the postulates concerning ${}_sE_R$ and V have been modified in the direction of greater complexity (in Hull's cyclostyled Memoranda on Behaviour Theory of Nov., 1949, Feb. 7th, 1950, March 6th, 1950). But these changes are not such as to invalidate the above deductions.

V

References

1. Bartlett, F. C. (1939). *Study of Society*. London.
2. Berlyne, D. E. (1949). "Interest" as a psychological concept. *Brit. J. Psychol.*, **39**, 184–195.
3. Hull, C. L. (1943). *Principles of Behavior*. New York.
4. Hull, C. L. (1949). Stimulus-intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, **56**, 67–76.
5. Titchener, E. B. (1901). *Experimental Psychology*. New York. (Vol. I, Part II, p. 207.)

(*Received 1st October*, 1949.)

THE IMAGINARY QUESTIONNAIRE

BY

J. W. WHITFIELD

(*From the Medical Research Council Applied Psychology Research Unit, Cambridge*)

The experiment was designed to throw some light on the statistical problems in the analysis of questionnaire data. Previous work (unpublished) suggested that a simple choice response was partially determined by previous responses; and also that the nature of the determination was changed with changing length of series. A "null" experiment was devised in the form of a questionnaire without any questions, and the distribution of responses was studied with respect to the problems formulated.

The observations are discussed in three sections.

In the statistical discussion an alternative meaning to overall association or dissociation is advanced. This relates association or dissociation to human behaviour in the serial response situation, rather than to qualities of the questionnaire. It is further suggested that association between specific questions should be tested against the association in the whole questionnaire, and an appropriate treatment is indicated.

The observations depart from statistical randomness in certain ways. Answers made up almost entirely of one form of response are given less often than would be expected. Long sequences of the same type of response are relatively infrequent, and sequences of alternation of response are also rare. As the material is "null" it implies that the human concept of randomness differs from the statistical concept.

An attempt is made to define the human concept of randomness. It appears that a series of responses which has a pattern, or for which the subject can postulate a simple "cause" will not be accepted as random by the human subject. This raises problems of a perceptual and cognitive nature. It also has a bearing on the design of questionnaires or experiments involving serial responses.

I

Aim of Experiment

The purpose of this experiment was to throw some light on the statistical problems involved in the treatment of questionnaire data. With purely objective "factual" types of questionnaire many of the problems discussed should not arise, but in those cases where the questions present a judgment problem to the subject, certain effects due to the nature of the presentation may be observed. Questionnaires in the field of public opinion, or those which purport to give evidence in the study of personality or temperament, and laboratory-type discrimination tasks come into this category.

Preliminary observations had led to two hypotheses:

(1) In a series of questions involving judgment with only two possible responses (e.g. Yes or No), a particular response is partially determined by previous responses.

(2) Length of series modifies the nature of the effect.

An examination of the statistical concept of independence necessary for the evaluation of the significance of any observations follows naturally from these hypotheses.

II

Material

As an attempt to produce a psychologically "null" condition, an Imaginary Questionnaire was produced—i.e., a questionnaire with space for questions and for a "Yes" or "No" response, but with no questions. The subjects were instructed to ask themselves questions of the form "Do you prefer .... to ....?" and record their answers in the appropriate column for "Yes" or "No." Three groups of subjects were used, the bulk of them being students of an elementary pyschology course, a few were more sophisticated in psychological experimentation. Three sizes of questionnaire were used, 6, 10 and 14 questions. It may be argued that this did not provide a null condition in the psychological sense, a point which will be discussed later, but it is as null as anything the experimenter could think of. The really important feature is that within the limits of two responses any further constraints on behaviour are provided by the subjects themselves.

III

Observations and Discussion of Statistical Problems

TABLE I

Distribution of Responses by Questions

Question number	*Six questions*		*Ten questions*		*Fourteen questions*	
	Yes	*No*	*Yes*	*No*	*Yes*	*No*
1	34	6	35	5	33	7
2	29	11	27	13	30	10
3	25	15	21	19	28	12
4	28	12	29	11	26	14
5	30	10	28	12	24	16
6	29	11	26	14	27	13
7	—	—	23	17	25	15
8	—	—	24	16	22	18
9	—	—	24	16	21	19
10	—	—	20	20	29	11
11	—	—	—	—	27	13
12	—	—	—	—	23	17
13	—	—	—	—	30	10
14	—	—	—	—	29	11

It will be seen in Table I that there are far more "Yeses" than "Noes." This may be due to the normal verbal ordering of the words—Yes, No, and never No, Yes. The effect is most marked with the first "question," where a Yes answer is obtained significantly more often than a No, even taking the general tendency into consideration. The verbal arrangement hypothesis is somewhat strengthened by observations on a small number of subjects who were asked "mentally" to toss up a coin, and record their guesses as to heads or tails. These showed a preponderance of "Head" responses on the first imaginary toss.

Overall Association or Dissociation and the Problem of Independence.

Usually, in the treatment of questionnaire data we wish to discover whether there is any association between the questions, or, specifically, if two or more of

them are associated in any way. There are $\frac{1}{2}n\ (n-1)$ pairs of questions, where n is the number of questions, and the same number of 2×2 contingency tables can be constructed. Slater (1947) has implied that these are independent contingencies, and has tentatively suggested that the relevant test of overall association in the questionnaire is the summated χ^2, from all the 2×2 tables with $\{\frac{1}{2}n\ (n-1)-1\}$ degrees of freedom. In my opinion this is erroneous, as the association between questions 1 and 3 is partially determined by the association, if it exists, between questions 1 and 2, and between questions 2 and 3 (see note on p. 80). We can, however, with safety summate the χ^2 for the $n-1$ independent 2×2 tables concerning the association of one question with each of the others, and test with $(n-1)$ degrees of freedom.

The problem of an overall association, treated as an exercise in factorial analysis by Slater in the same paper can be considered along other lines. Taking the observations from the Imaginary Questionnaires and constructing the matrices of 2×2 tables we find the following (Table II):

TABLE II

Number of questions	*Total observed frequency in Yes-Yes (YY) cells*	*Total expected frequency in YY cells (calculated in the usual manner assuming independence)*
6	312	318·475
10	729	740·900
14	1,638	1,621·650

In a 2×2 table relating a pair of questions, if there are more YY than expected it indicates association between the questions, if less it implies dissociation. Table II summates the observations for all pairs of questions.

Thus there is evidence of slight overall dissociation in the 6- and 10-question series. A first general factor of dissociation seems to the writer to be a difficult concept, and an alternative approach is desirable.

If we consider the distribution of responses amongst the subjects we see reasons for these peculiar effects. The total number of YY entries is a function of the distribution of Y amongst individuals. For example, if a subject has said Yes to five questions he will have contributed 10 YY, as five questions make up 10 question pairs. Similarly, if he has said Yes to six questions he will have contributed 15 YY, and so on. Unless the distribution of Y amongst the individuals corresponds to that expected on independence, discrepancies of the nature demonstrated are bound to happen. The distribution on independence assumes that there is a constant probability of giving a Y to a specific question, and that different individuals are merely different trials, with the same probability. The probability of a Y differs between questions, and the observed proportion of Y to a particular question is used as the Y probability for that question. Thus the distribution of Y in trials (i.e. in individuals) is the expansion of p_1, p_2, p_3, etc., where p_1, p_2, p_3, etc., are the probabilities of a Y response to questions 1, 2, 3, etc., as derived from the data. The relationship in the questionnaire may be considered, therefore, as partly composed of the true association between questions, and partly of the behaviour of individuals in answering questionnaires.

TABLE III

DISTRIBUTION OF YESES AMONGST INDIVIDUALS

No. of Yeses	*Six questions*		*Ten questions*		*Fourteen questions*	
	Obs. f.	*Exp. f.*	*Obs. f.*	*Exp. f.*	*Obs. f.*	*Exp. f.*
0	0	0·01	0	0·00	0	0·00
1	1	0·23	1	0·02	0	0·00
2	0	1·65	0	0·16	0	0·00
3	4	6·16	0	0·86	0	0·02
4	15	12·62	0	2·89	1	0·12
5	18	13·47	5	6·51	0	0·49
6	2	5·86	13	9·90	1	1·54
7	—	—	16	10·10	2	3·63
8	—	—	4	6·61	11	6·46
9	—	—	1	2·50	10	8·67
10	—	—	0	0·40	1	8·70
11	—	—	—	—	9	6·27
12	—	—	—	—	3	3·08
13	—	—	—	—	1	0·92
14	—	—	—	—	1	0·13
Obs. variance ..	0·834		1·597		3·828	
Exp. variance ..	1·158		2·189		3·01	
χ_2	5·670		8·445		12·013	
Degrees of freedom	2		2		3	
P	about 0·05		<0·02		<0·01	

Total $\chi^2 = 25{\cdot}128$, 7 d.f. $P = <0{\cdot}001$.

The labour of expansion of P_1, P_2, P_3, etc., can be reduced—

$P_0 = q_1 . q_2 . q_3 . \ldots\ldots q_n$.
$P_1 = P_0 \ \Sigma(p/q)$
$P_2 = {}^1/_2 \{P_1 \ \Sigma(p/q) - P_0 \ \Sigma(p/q)^2\}$
$P_3 = {}^1/_3 \{P_2 \ \Sigma(p/q) - P_1 \ \Sigma(p/q)^2 + P_0 \ \Sigma(p/q)^3\}$
$P_n = {}^1/_n \{P_{n-1} \ \Sigma(p/q) - P_{n-2} \ \Sigma(p/q)^2 + P_{n-3} \ \Sigma(p/q)^3 -, +, -, \ldots (-1)^{n-1} P_0 \ \Sigma(p/q)^n\}$

P_0, P_1 etc. = probability of an individual giving 0, 1, etc. Yes responses.

Testing the Association Between Specific Questions.

It will be seen from Table III that the expected distribution on independence is not a satisfactory fit. Thus we can say that independence does not hold. We shall consider later why this should be so. The practical problem lies in the 2 × 2 table, relating two particular questions, and how to calculate appropriate expected frequencies in the cells. To assume independence in the normal way may be misleading, if the question we really wish to pose is "Are these two questions more associated or less associated than other question pairs in the questionnaire?" This question can be considered by distributing the observed number of agreements (i.e. *YY* and *NN*) equiprobably amongst the question pairs. Each question pair will therefore be given expected frequencies of *YY* and *NN* (and *YN* and *NY*) which are functions of the Yes and No frequencies for the questions and also of the overall association or dissociation in the questionnaire.

The total observed agreement between questions (ΣO) is

$$\Sigma \tfrac{1}{2} a\,(a-1) + \Sigma \tfrac{1}{2}\,(n-a)\,(n-a-1)$$

where a = number of responses by an individual
n = number of questions.

This becomes

$$\tfrac{1}{2}\,mn\,(n-1) - nY + \Sigma a^2$$

where m = number of individuals
Y = total number of Y responses.

The total agreement expected on independence (ΣE) is

$$1/m\,\Sigma\,\{y_1.y_2 + (m-y_1)\,(m-y_2)\}$$

where $y_1.y_2$, etc., are the number of Y responses to questions 1, 2, etc.
This becomes

$$1/m\,\{Y^2 - \Sigma y^2 + \tfrac{1}{2} n\,(n-1)m^2 - m\,(n-1)Y\}$$

Unless the distributions of Y amongst individuals corresponds to that expected on the independence hypothesis there will be a discrepancy $\Sigma O - \Sigma E$. In order to test the association between a specific pair of questions i and j, this discrepancy will have to be distributed amongst all the question pairs. The agreement expected on complete independence in the specific pair is the expected frequency of YY (i.e., $(y_i.y_j)/m$) plus the expected frequency of NN $((m-y_i)\,(m-y_j)/m)$. The required expectations will differ by an amount "z" from these, "z" being related to $\Sigma O - \Sigma E$.

Before defining "z" it is worthwhile to note in passing that $\Sigma O - \Sigma E$ is related to the quantity S in Kendall's coefficient of rank correlation. It has been shown that S in the 2×2 table is equal to the difference between the cross products, and that the denominator required for τ is $\sqrt{A.a.B.b}$, where A,a,B,b are the marginal totals (Whitfield, 1947). If we were to calculate S for all possible 2×2 tables in the questionnaire we would find that $S = m\,(\Sigma O - \Sigma E)/2$. The total S′ (denominator) is $\tfrac{1}{2}\{(\Sigma\sqrt{b})^2 - \Sigma b\}$ where b_1, b_2, etc., are $y_1(m-y_1)$, $y_2(m-y_2)$, etc. Thus, if we wished, we could express the overall association or dissociation as a τ correlation

$$\tau = \frac{m\,(\Sigma O - \Sigma E)}{(\Sigma\sqrt{b})^2 - \Sigma b}$$

but as this is mixed up inextricably in a lot of partial relationships it does not appear to be capable of a significance test directly.*

Using this information, however, we can calculate the quantity "z."

$$z_{ij} = \frac{b_i b_j\,(\Sigma O - \Sigma E)}{(\Sigma\sqrt{b})^2 - \Sigma b}$$

To consider an example. In the 10-question questionnaire questions 5 and 7 appear to be related.

* It is likely that with moderately large numbers both of observers and questions the value $(m-1)\,(m\Sigma a^2 - Y^2)/(mY - \Sigma y^2)$ will be distributed as χ^2 with $(m-1)$ degrees of freedom. It will be necessary to discover more about the conditions under which this approximation is reliable as a significant test. It is derived from Kendall. "Rank Correlation Methods"—on the problem of "m" rankings, regarding the questions as ranking the observers into two sets of tied ranks.

		Question 7 Y	N	
	Y	12	16	28
Question 5	N	11	1	12
		23	17	40

The frequency of *NN* expected on complete independence is 5·1 χ^2 with Yates' correction is 6·313, which suggests a markedly significant degree of dissociation. But there is evidence in the 10-question questionnaire of overall dissociation, and we therefore wish to see whether questions 5 and 7 are more dissociated than the average.

$\Sigma O - \Sigma E$ is $-23{\cdot}8$

Question	b	$\sqrt{b}$
1	175	13·23
2	351	18·73
3	399	19·97
4	319	17·86
5	336	18·33
6	364	19·08
7	391	19·77
8	384	19·60
9	384	19·60
10	400	20·00
	3,503	186·17

$$\text{Overall } r \quad \frac{-23{\cdot}8 \times 40}{186{\cdot}17^2 - 3{,}503} = -0{\cdot}031$$

$$z_{57} \text{ is } -\frac{23{\cdot}8 \times 18{\cdot}33 \times 19{\cdot}77}{186{\cdot}17^2 - 3{,}503} = -0{\cdot}28$$

Thus the new expected *NN* frequency is 5·10 − 0·28 = 4·82. This reduces χ^2 to 5·298. Against this must be set the fact that we have reduced the degree of freedom, presumably to 8/9. Thus there is still evidence that the two questions are dissociated, whatever that may mean in the imaginary case.

IV

Behavioural Observations in the Imaginary Questionnaire

Certain aspects of the subjects' behaviour have already been referred to in relation to the statistical treatment. Along with other features they must further be considered in describing the human "random" performance.

Avoidance of Extremes.

The reader will have noted in Table III that, with one exception, fewer individuals than would be expected gave a large preponderance of one response. (The exception is in the case of the Yes response to the 14-question questionnaire.) This accounts for the reduced variance of the 6- and 10-question series. Some previous experiments

had suggested that the relation of observed to expected variance was affected by the number of questions presented. A short series gave a variance much smaller than the expected, but a long series reversed the effect. This is substantiated by these later observations.

The case of the reduced variance in short series can perhaps be explained by arguing that a marked preponderance of one response is not compatible with the subject's idea of random behaviour—it suggests a "cause," i.e. there must be a reason for departure from an average distribution. If this is true, and there is further evidence which appears to support the idea, then it is not surprising that a 6—0 distribution of responses is avoided, constraining the responses into a more limited range. It is also likely that the amount of imbalance of frequency of responses which will be tolerated by the subject is a proportion of the total available range. Hence the tolerated range will increase linearly with increase in the number of questions. The standard deviation of the expected distribution, however, is increased at the rate of approximately the square root of the number of questions, and therefore the relative absence of individuals with marked unbalanced responses becomes of progressively less effect on the variance of the distribution. There may also be a factor working in the opposite direction. If some individuals are more prone to give Yes responses, and some prone to give No responses, even though both may keep within the "tolerated" range there is, as it were, enough room in the longer series for them to demonstrate their proneness, and hence increase the variance.

The Effect of a Response on Subsequent Responses.

Another approach may be made to the question of the difference between the human idea of "random" and true randomness. The question may be asked: "What effect, if any, has a response on its successor, or later responses?" Lag correlations were calculated (i.e., 1st with 2nd, 2nd with 3rd, etc., then increasing the interval—1st with 3rd, 2nd with 4th, etc., and so on). These are shown in Figures 1, 2 and 3. If there were no series effects—i.e. if the probability of a Yes or No response was completely unaffected by previous responses, we would have correlograms which were straight lines at zero, if there were no differences between subjects, or at a positive value if there were a distribution among subjects of Yes-proneness or No-proneness. If, on the other hand, a Yes response inevitably produced a No response as its immediate successor, and vice versa, the correlogram would oscillate — 1·0, + 1·0, — 1·0, etc.

CORRELOGRAMS

FIGURE 1. SIX QUESTIONS

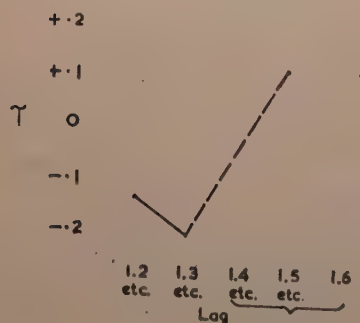


FIGURE 2. TEN QUESTIONS

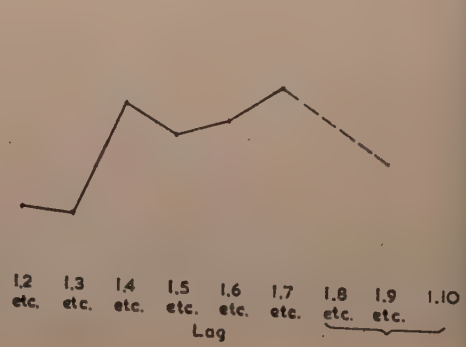
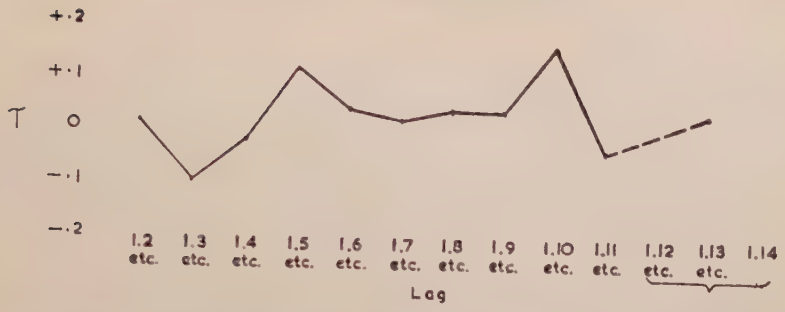


FIGURE 3. FOURTEEN QUESTIONS



It is possible to make some reasonable conjectures as to the behaviour of the subjects from some of the features of the correlogram. Let us consider the 10-question case—the maximum positive correlation occurs with the relations, Questions 1 and 4, Questions 2 and 5, etc. This suggests a series with sequences of not greater than two same responses. A chance series obeying this—i.e. all possible series which do not infringe the limit of two on sequences—gives a correlogram shown in Figure 4. It differes from the observed correlogram in two main respects—(*a*) more extreme values are obtained, and (*b*) the difference between the 1·2 type correlation and the 1·3 type correlation is not the same as that observed. A "dilution" of the series—i.e. allowing a few series with sequences longer than two members may account for point (*a*), and the difference (*b*) may be accounted for by yet another form of avoidance by the subjects. The lower observed correlation of the 1·3 type implies that subjects have avoided an alternation pattern in their responses. This again fits in with the general idea of avoiding apparent "cause." A chance series, with 10 questions, allowing no sequences greater than 3, up to

FIGURE 4.

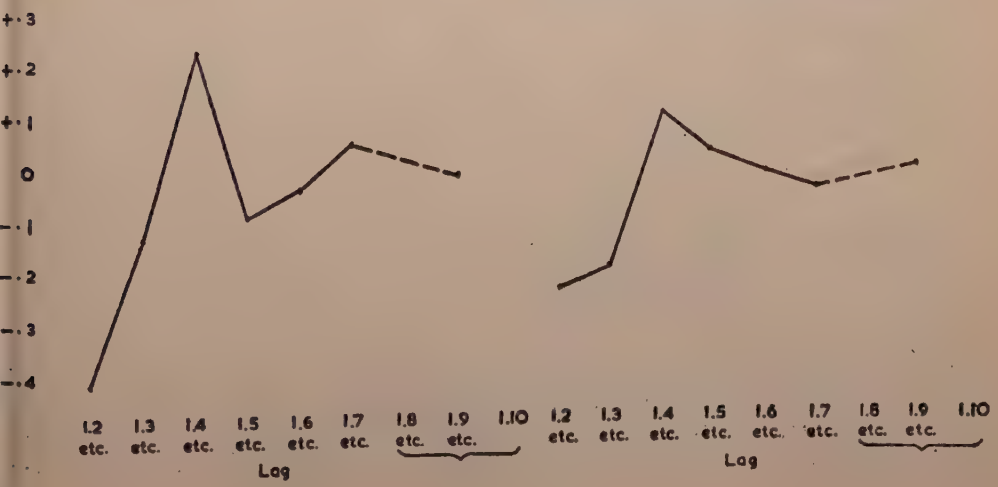
TEN QUESTIONS—CHANCE DISTRIBUTION A

All series within limits of no sequence of same response greater than two.

FIGURE 5.

TEN QUESTIONS—CHANCE DISTRIBUTION B

All series within limits of no sequence greater than three, not more than one sequence of three, and no alterations *YNYNYNY* or converse or greater.



one sequence of 3, and avoiding alternation sequences of *YNYNYNY*, or converse, or greater, gives the correlogram shown in Figure 5, which in the points considered is moderately satisfactory.

The 14-question series shows a somewhat different pattern, but similar in its main essence—slightly longer sequences would appear to be tolerated, and alternation again avoided. It is also worthy of note that the mid sample of 10 questions from the total of 14 gives a correlogram similar to that of the whole 14, and not to that of the 10 questions, suggesting that the effect is due to the displayed length of series, and cannot be attributed to any marked change of behaviour at the ends of the questionnaire.

The hypothesis concerning the tolerated sequence length may be tested in another way. Given an individual who has shown, let us say, 6 Yes responses to 10 questions —these 6 Yeses may theoretically be distributed anywhere in the 10 questions.

TABLE IV

SEQUENCE LENGTH

	No. of Yeses	*No. of cases*	*No. with sequence of Yeses* $> \frac{1}{2}$ *Yes frequency*	*Probability of as few such cases or fewer*
14 questions ..	14	1	1	No chance involved
	13	1	1	" " "
	12	3	1	0·2263
	11	9	2	0·01965
	10	1	0	0·6510
	9	10	4	0·6871
	8	11	0	0·09076
	7	2	0	0·5184
	6	1	0	0·8660
	4	1	0	0·8790
10 questions ..	9	1	1	No chance involved
	8	4	1	0·1111
	7	16	7	0·04997
	6	13	3	0·002982
	5	5	1	0·1875
	1	1	1	No chance involved
6 questions ..	6	2	2	No chance involved
	5	18	18	" " "
	4	15	7	0·2231
	3	4	4	No chance involved
	1	1	1	" " "

Taking into consideration only those cases in which there is a chance of obtaining different frequencies, and employing $\chi^2 = -2\Sigma \log. p$, with 26 degrees of freedom, χ^2 is 48·339, and the combined probability is less than 0·01.

They can be considered as being distributed by the Noes into 5 sequences—first sequence between the beginning of the questionnaire and the first No, second sequence between the 1st and 2nd No, 3rd between 2nd and 3rd No, 4th between 3rd and 4th No and 5th between the 4th No and the end of the questionnaire. The

chance distribution of six objects into five categories is easily calculated. In the lower numerical values of sequences an individual may contribute more than once, and so as a test individuals were divided into two classes—those who gave a sequence of YY longer than half their total YY, and those who did not. The chance of a random series, with the same number of YY showing a sequence greater than half the total Yes frequency was then calculated, and the comparison shown in Table IV. It will be seen that subjects tended to avoid sequences of such length. Unfortunately, a test of the avoidance of alternations would be, at the least, very tedious.

Summing up the evidence we can therefore make the following reasonable conjectures about behaviour in the Imaginary Questionnaire:

(1) There is a marked tendency to answer Yes to the first question.

(2) The distribution of Yeses amongst individuals is not chance. It is—

  (*a*) Leptokurtic with short series;

  (*b*) Less certainly—it is platykurtic with longer series, with a suggestion that the change-over takes place somewhere about the 12-question length.

(3) Following from (2) even if the purely statistical difficulties of independence in a matrix of 2×2 tables about the questionnaire as a whole are overcome, expected cell frequencies cannot be calculated in the usual manner from the marginal totals, assuming an independence between questions.

(4) There exist within the series of responses sequential effects which can be described as the avoidance of long sequences of the same response, and also the avoidance of sequences of alternation.

V

Discussion of the Psychological Problem

Any statement beyond those purely descriptive of findings is, of course, pure conjecture. A few rather wide hypotheses were formulated before these reported experiments were begun. These have been modified, but still remain unverified, and to a large extent difficult to verify. Nevertheless, they may exist as cautions in psychological experimentation. Nature is said to abhor a vacuum—the human subject seems to do so, and objects to thinking about a problem which has insufficient data. Introspective reports from Ebbinghaus-type memory experiments suggest that subjects often attempt to make some sort of "sense" out of the "nonsense" syllables employed. Bartlett's experiments on social thinking show that even when a considerable amount of information is presented to the subject the normal modes of solution call for the introduction of generalisations or other information not presented in the data and brought in by the subject himself In the case of the Imaginary Questionnaire the generalization introduced by the subject is that the responses must not imply any consistent "cause." There is no reason that he should show a relationship between the imaginary questions. The behaviour must then conform to the subject's idea of relationless, i.e. random behaviour. Several subjects reported that in some cases they had to reverse the form of the question they were asking themselves to avoid repeating the same response more often than they thought desirable. If the observations had shown longer sequences than expected they would have shown consistency of set towards Yes or No in the choice

of questions the subjects produced for themselves—which would perhaps be understandable. The fact that the reverse is true, taken together with the subjects' reports shows that there was an attempt made, consciously by some though probably not all, to randomize the responses. That the responses are not distributed in a statistically random fashion implies a difference between the statistical concept of randomness and the human concept of random behaviour.

The usual human idea of randomness of a series of events is one which shows no apparent cause. This is primarily a perceptual problem. One is unlikely to be dealt four Aces, four Kings, four Queens and the Jack of Spades in a card game, but the same unlikelihood exists for any other specific hand. The former would be perceived and a cause (bad shuffling, good luck, etc.) would be invoked to account for it, whereas in the second case nothing would be thought or said—unless the hand had been specified beforehand. The early writers on probability were concerned with the relation of probability to belief. The question—how often must a penny come down heads in succession from the first before the observer may reasonably doubt its being a perfect penny?—is one which has no generally acceptable statistical solution, even if information is supplied as to the relative frequency of double-headed pennies in circulation. A considerable amount of attention has been given to problems of this nature (see Keynes (1921) for a full discussion). But, except under unlikely condition of being constrained to act in an absolutely logical manner, the usual human response is to make some judgment. After a sequence of events has shown a relationship consistent with a "cause," the observer is inclined to accept the "cause" as demonstrated, or at least to believe he will be more often right than wrong if he so decides. The odd-looking concatination of events, which would happen if unrelated once in a hundred times is considered as if it were impossible of occurrence by chance, but only by design.

The Imaginary Questionnaire behaviour is in a sense the converse of this; no cause means that there should be no series of events which give the impression of "cause." It would appear that "cause" is attributed to a series which possesses a descriptive simplification, or other distinguishing characteristic. Three such "causes" tended to be avoided by subjects:

(1) Preponderance of one response over another, which would be taken to imply a mechanism giving preference to one response rather than the other. As no evidence of such a mechanism is otherwise present, this is avoided.

(2) "Cause" in the sense of long sequences. There is by and large a practical belief in a "law of averages"—or, at a more sophisticated level, a treatment of chance processes as if they were derived from finite and limited series.

(3) Alternation suggests a simple mechanism of inhibition, or refractory period. Again, failing other evidence in the nature of the experiment to suggest the existence of such a mechanism, the subject avoids alternative responses.

Although it is of interest to know something about the human concept of random behaviour, it is perhaps the relevance of this to the more practical field of experimentation which is important. In experimental design the question of randomization crops up, and the difference between the random tables and the human concept may be important. This is particularly true in those cases where the task is difficult either by absence of sufficient data—as in threshold measurements, or by working near the limits of performance—as in certain forms of intelligence test, or by difficulty of appreciation or apparent irrelevancy—as in many questionnaires used in work on temperament. It would seem desirable to randomize, where possible, in a humanly

acceptable way. An ad hoc solution would be to randomize in short sequences, probably of about 10, giving equal frequencies to correct responses or expected responses. There is also the possibility that the effect could be used directly in certain types of problem. Under what conditions is a subject prepared to behave differently from his concept of randomness in a series which he believes to be truly random? It is possible that this could lead to an objective estimate of certainty of response.

VI

References

1. Bartlett, F. C. (1938). "The co-operation of social groups." *Occup. Psychol.*, **12**, 30–42.
2. Kendall, M. G. (1948). *Rank Correlation Methods.* London.
3. Keynes, J. M. (1921). *A Treatise on Probability.* London.
4. Slater, P. (1947). "The Factor Analysis of a Matrix of 2 × 2 Tables." *Sup. J.R.S.S.*, **18**, 114–127.
5. Whitfield, J. W. (1947). "Rank Correlation between two Variables, one of which is Ranked, the other Dichotomous." *Biometrika*, **34**, 292–296.

(*Manuscript received* 29th *November*, 1949.)

ON THE RELATION BETWEEN TRANSFER AND DIFFICULTY OF INITIAL TASK

BY

J. SZAFRAN and A. T. WELFORD*

(From the Nuffield Research Unit into Problems of Ageing, the Psychological Laboratory, Cambridge)

Subjects were required to throw at a target under three conditions of varying difficulty. Six groups of 14 subjects were used, each group receiving one of the six possible orders of presentation of the three conditions.

Results in terms of mean distance from the "bull" indicated that the three conditions were producing different transfer effects. These could not be adequately explained in terms of stimulus- or response-similarity, and accordingly a tentative hypothesis is put forward in terms of the relative difficulty of the condition presented first, to those presented subsequently. It is suggested that transfer tends to be positive from a relatively difficult initial task to a subsequent task which is easier, while transfer will tend to be negative from a relatively easy initial task to one which is more difficult.

The first of the three conditions presented exerted considerably more powerful transfer effects than the second. Three possibilities as to why this should be so, are outlined.

I

Introduction

Transfer of training has usually been treated as a matter of similarity between two tasks as regards either the stimulus presented or the response required or some relation between them, and experiments on transfer have usually taken the form of presenting two tasks in which the second differs from the first in one or other of these respects, and assessing the gain or loss in performance of the second task resulting from performance of the first. "Stimulus" and "response" have in these studies been given connotations of varying width from nonsense syllables and digits to general principles of complex displays and methods of work. But whether they have been detailed or general, the process of transfer has usually been conceived within a framework in fact, if not in name, rigidly associationist, and typified by the use of the term "identical elements" for what is transferred.

It seems clear, however, that this traditional approach deals with a type of carry-over from one task to another which is only a special case of the more general principle that performance at one task affects performance at subsequent tasks in many different ways. It is the purpose of this paper to describe some experimental results indicating another type of carry-over which can be subsumed under the general heading of transfer, but which lies outside the scope of the subject as usually treated.

II

Method

The subjects' task was to throw loops of light chain about 3 inches long at a target consisting of intersecting strips of fibre board arranged to form 7 × 7 3-inch square "boxes," the centre box is being outlined in white to act as the "bull."

* This paper is based on work an account of which was given in N.R.U.P.A. Report No. 3, by J. Szafran, dated June, 1948.

This design of target was chosen not only for the practical reason that the chains fall into it "dead" and do not slide after they have landed, but because it enables the separate scoring of inaccuracies on the "far-near" and "left-right" dimensions. Far-near inaccuracy not only differs from left-right owing to the foreshortening of this dimension as seen by the subject, but would seem to be of different origin. In throwing underhand, for instance, inaccuracy on the left-right dimension is due essentially to a failure to throw in the correct direction, whereas that on the far-near dimension is due either to an error of "elevation" or to releasing the chain at the wrong time, or to throwing too hard or not hard enough, or to some combination of these. The design of target was justified by results as the inaccuracies on the two dimensions behaved very differently in some respects. For our present purposes, however, they were very similar, and in consequence we have for the sake of simplicity added the inaccuracies on the dimensions together when presenting results.

Each subject threw 50 chains under each of three conditions:—

A. The subject threw directly at the target which was placed flat on the floor some 8 feet away.

B. The same as *A*, except that the subject was required to throw over a horizontal bar 32 inches from the ground and 5½ feet away. The target was in full view underneath the bar so that as regards display the task was unchanged. The effector task was, however, made more difficult by the constraint that the subject had to throw in a high "lob" to clear the bar.

C. The subject threw over a screen which hid the target from direct vision so that it could only be seen *via* a mirror placed beyond it. As the height and distance of the screen were the same as those of the bar, the effector task was the same as that of *B*. The screen and mirror, however, made the perceptual task more difficult, especially as regards aiming in the far-near dimension which was reversed in the mirror.

The 84 subjects, who came from Polish Resettlement Corps Camps and Hostels in Cambridgeshire, were divided into six equal groups each of which was given one of the six possible orders of presentation of the three tasks. Assignment to these groups was at random, except that they were balanced as regards age.

The subjects were instructed to aim for the 3-inch square which formed the centre of the target and to take their own time. They threw the whole 150 chains at a single session with short breaks between the groups of 50 for the recording of results.

III

Results

The group average results for the three tasks are set out in Table I, from the last two columns of which it will be seen that there are small but consistent differences

TABLE I

Mean Inaccuracy in inches per throw for the six Groups of Subjects having Different Orders of Presentation of three Conditions of Throwing

Order of Presentation	*Condition Presented*			*Mean of Group*	*Mean of two Groups having same condition First*
	First	*Second*	*Third*		
Direct, Bar, Screen ..	5·49	5·08	6·96	5·84	
Direct, Screen, Bar ..	5·68	7·22	5·19	6·03	5·93
Bar, Direct, Screen ..	5·27	4·79	7·31	5·79	
Bar, Screen, Direct ..	5·32	7·10	4·57	5·66	5·73
Screen, Direct, Bar ..	7·13	4·20	4·74	5·35	
Screen, Bar, Direct ..	6·86	4·79	4·41	5·35	5·35
Means	5·96	5·53	5·53	—	—

between the various order of presentation groups. The average inaccuracy of the groups who first threw direct is higher than that of the groups who first threw over the bar, and these again are higher than those who first threw over the screen.

This is a surprising result not to be expected from an experiment of this design, which is essentially an expanded form of the well-known transfer paradigm:

Matched Groups	*First Task*	*Second Task*
Group 1	A	B
Group 2	B	A

With this design, unequal difficulty of the two tasks will, if it can be represented in the scores as a difference of simple additive factors, make no difference to the totals of the two groups for both tasks together. If, on the other hand, the inequality of difficulty is such as to require representation in the scores by different multiplicative factors, the combined effects of difficulty and improvement with practice will cause the total for the group which receives the more difficult task first to be higher than that for the group which receives the easier first—the opposite of the present finding.

Unless some very remarkable mischance of sampling has occurred, it seems clear that there is an interaction between task and position in the order of presentation. This is brought out by the figures in Table II, from which it will be seen that while inaccuracy in the direct condition was substantially lower when this condition was presented second or third than when it came first, in the screen condition it was a little higher in the second and third places than in the first.

TABLE II

MEAN INACCURACY IN INCHES PER THROW FOR THREE CONDITIONS OF THROWING

Condition	*Presented First*	*Presented Second*	*Presented Third*	*All three Positions*
Direct	5·59	4·50	4·49	4·86
Bar	5·29	4·94	4·96	5·06
Screen	6·99	7·16	7·13	7·09

Two possible explanations of this interaction appear to deserve consideration:

(*a*) that the three conditions are not equally amenable to practice—i.e. practice effects are greater in direct throwing than in the other two conditions,

(*b*) that the different tasks produce different transfer effects.

It is impossible to make any adequate assessment of practice effects from the present experiment, but some indication of them for the direct and screen conditions can be obtained from the results of nine additional subjects who threw two separate sets of 50 chains direct, and of 52 additional subjects who similarly threw twice over the screen. The improvements when the second sets were compared with the first were some 7 and 12 per cent. (significant) respectively—i.e. improvement in direct throwing was *less* than in throwing over the screen. This is the opposite of what would be expected if the interaction observed in the present experiment were due to the conditions not being equally amenable to practice. It is true that the practice effect for direct throwing is assessed from the results of only a few subjects and is therefore not very reliable, but even if these are excluded from consideration altogether, the practice effect for the screen condition is greatly at variance with the figures for this condition shown in Table II. It seems, therefore,

that an explanation of our results in terms of different transfer effects produced by the three conditions of throwing is to be preferred to an explanation in terms of differing practice effects. We will proceed to examine these transfer effects in greater detail.

Transfer Between First and Second Tasks.

Comparing the means for the groups of 14 subjects who received each task after each of the others, with the means for the 28 subjects who received the same task first, we obtain figures for absolute and percentage transfer as set out in Table III. The noteworthy points about this table are the substantial positive transfer effects from the bar and screen conditions to the direct condition, and the small *negative* transfer effects from the direct and bar conditions to the screen condition.

TABLE III

TRANSFER EFFECT IN INCHES PER SUBJECT PER CHAIN FROM FIRST TASK TO SECOND

Presented First	*Presented Second*	*Transfer Effect*	
Direct	Bar	+ 0·21	+ 4%
Direct	Screen	− 0·22	− 3%
Bar	Direct	+ 0·79 *	+ 14%
Bar	Screen	− 0·11	− 2%
Screen	Direct	+ 1·39 *	+ 25%
Screen	Bar	+ 0·50	+ 10%

* Indicates significance at the 1% level.

The transfer effect shown in the third column is the difference between the mean for the task concerned of the 14 subjects who received the task in the second position, and the mean for the task concerned of the 28 subjects who received it in the first position.

Importance of the Task Presented First.

TABLE IV

TRANSFER EFFECT IN INCHES PER SUBJECT PER CHAIN FROM FIRST TASK TO THIRD

Presented First	*Presented Third*	*Transfer Effect*	
Direct	Bar	+ 0·11	+ 2%
Direct	Screen	+ 0·04	+ 1%
Bar	Direct	+ 1·01 *	+ 18%
Bar	Screen	− 0·31	− 5%
Screen	Direct	+ 1·17 *	+ 21%
Screen	Bar	+ 0·56 †	+ 10%

* Indicates significance at the 1% level.
† Indicates significance at the 5% level when tasks presented second and third are considered together.

The transfer effect shown in the third column is the difference between the mean for the task concerned of the 14 subjects who received the task in the third position, and the mean for the task concerned of the 28 subjects who received it in the first position.

The indication from the last column of Table I that transfer effects are dependent upon which task is presented as the first of the three is confirmed by Tables IV and V. From Table IV it will be seen that the transfer effects of the task presented first on the task presented third, are essentially similar to those on the task presented second, while from Table V it will be seen that the transfer from the second to the third tasks is in all cases small.

TABLE V

TRANSFER EFFECTS IN INCHES PER SUBJECT PER CHAIN FROM SECOND TASK TO THIRD

Presented Second	*Presented Third*	*Transfer effect when subjects receiving task presented third are compared with all subjects receiving that task second.*		*Transfer effect when subjects receiving task presented third are compared with those subjects who received that task after the same first task.*	
Direct	Bar	+ 0·20	+ 4%	+ 0·05	+ 1%
Direct	Screen	− 0·15	− 2%	− 0·20	− 3%
Bar	Direct	+ 0·08	+ 2%	− 0·21	− 5%
Bar	Screen	+ 0·19	+ 3%	+ 0·26	+ 4%
Screen	Direct	+ 0·08	+ 2%	+ 0·22	+ 5%
Screen	Bar	− 0·25	− 5%	+ 0·11	+ 2%

The transfer effect shown in the third column is the difference between the mean for the task concerned of the 14 subjects who received the task in the third position, and the mean for the task concerned of the 28 subjects who received it in the second position. The effect shown in the fifth column is the difference between the mean for the same 14 subjects and 14 subjects who received the task concerned in the second position, both groups having received the same first task.

None of the values are significant.

IV

DISCUSSION

The Cause of the Observed Transfer Effects.

It does not seem possible to account for the transfer effects indicated in Tables III and IV in terms of either stimulus- or response-similarity. On the hypothesis that positive transfer occurs when the *stimulus presented* by the second task is similar to that of the first we should expect:

(*a*) A relatively large positive transfer from direct throwing to throwing over the bar, and vice versa, since the display in these two conditions was approximately the same.

(*b*) A relatively small positive or a negative transfer from these conditions to throwing over the screen, and vice versa, since the display in the screen condition was considerably different from that of the others.

With regard to (*a*), it will be seen from the tables that although the expected positive transfer occurs from the bar to the direct condition, similar transfer in the opposite direction is very small. With regard to (*b*), the requirements of the hypothesis are fulfilled by the negative transfer from both the direct and bar conditions to the screen condition, but are strikingly contradicted by the large positive transfer from the screen to the direct condition.

On the hypothesis that positive transfer occurs when the *response required* by the second task is similar to that of the first we should expect:

(*a*) A relatively large positive transfer from throwing over the bar to throwing over the screen, and vice versa, since the effector action required was the same in both conditions.

(*b*) A relatively small or negative transfer from direct throwing to both the other conditions, although not necessarily vice versa, because while the effector action likely to be adopted when throwing direct was inappropriate to the other conditions, that likely to be adopted in these other conditions could be used when throwing direct.

With regard to (*a*), it will be seen from Tables III and IV that while a positive transfer occurs when the bar condition is preceded by the screen condition, the transfer in the opposite direction is negative—i.e. the reverse of that required. With regard to (*b*), the expected transfers from direct throwing to the other conditions occur, but while the direction of transfers in the opposite direction are not inconsistent with the hypothesis, their magnitude relative to those discussed in (*a*) is out of all proportion to what would be predicted on the basis of response-similarity.

The inadequacy of both these hypotheses in the present case, leads us tentatively to suggest an alternative in terms of *relative difficulty* of the tasks. It is that—

(i) *When an easy task follows a difficult, transfer will tend to be positive.*

(ii) *When a difficult task follows an easy, transfer will tend to be negative.*

Predictions made on this hypothesis are, of course, subject to modification by the fact that some aspects of the two performances may be similar and lead to transfer along the lines of the conventional transfer theories. This applies especially to positive transfer appearing in the form of "practice effects," which would tend to increase (i) and diminish (ii), so that in a practical case an easy task following a difficult might show a large positive transfer effect while a difficult task following an easy might show a transfer effect which, while still positive, was smaller. The proviso should probably also be made that the difficult task must not be so difficult as to cause failure to achieve any reasonable standard of performance.

On this theory we should expect transfer to be:

(1) Substantially positive from the most difficult condition, i.e. throwing over the screen, to both the others.

(2) Substantially positive from throwing over the bar to direct throwing.

(3) Small or negative from direct throwing to both the other conditions.

(4) Small or negative from throwing over the bar to throwing over the screen.

It will be seen from Tables III and IV that the pattern of transfer effects observed is adequately accounted for by this theory.

When the scores are expressed in units which permit them to be regarded as the sums of additive factors, it is convenient from certain points of view to state this hypothesis in another way:

If two tasks of unequal difficulty are presented, achievement at the two tasks considered together, after allowance has been made for practice effects, will tend to be greater when the harder is presented first, than when the easier is presented first.

The hypothesis stated in this form agrees with the means in the last column of Table I, where, as has been previously indicated, the mean inaccuracy for all three tasks taken together was least when the subjects threw first in the hardest condition, i.e. over the screen, and greatest when they threw first in the easiest, i.e. direct.

The Nature of the Transfer Observed.

The large and consistent transfer effects from the first task to the second and third, and the small transfer effects from the second to the third, make it clear that the transfer with which we are dealing is not something which follows inevitably and automatically upon performance of any task, but is to be sought in terms of some influence of the *initial* task upon performance at subsequent tasks. As to the nature of this influence, three possibilities appear to deserve consideration:

(*a*) *The receptor comprehension and effector organization* necessary for the performance of a task is more easily built up than modified, so that a subject will tend to deal with subsequent tasks by means of the comprehension and organization built up during the performance of the first. A difficult task will produce a greater measure of comprehension and organization than will an easy, so that a subject who has a difficult first task will approach subsequent easy tasks with what is in a very real sense a greater ability than he would otherwise have had. When, on the other hand, an easy task comes first, the comprehension and organization it produces will be inadequate for dealing with subsequent difficult tasks, and will require modification when these are performed.

(*b*) The subject approaches the first task with some *expected standard of performance* which is preformed and more or less independent of the condition under which the task is presented, e.g. he regards a bull as "good" and an inner as "fair" and anything else as "poor." When he approaches a second task he recognizes that it is easier or more difficult than the first, and adjusts his standard of performance accordingly. The standard set for an easy task which follows a difficult will thus be higher than that set for the same task when it comes first. Similarly the standard set for a difficult task which follows an easy will be lower than that set for the same task coming first.

(*c*) The subject approaches the initial task with some expected standard of performance as described above, and adjusts his *level of care, effort*, etc., to attain this standard. When the initial task is difficult, this level will be high; when it is easy, the level will be low. If the initial level carries over to subsequent tasks, those which are easier than the initial task will be done better than they would have been had they come first in the series, while those which are more difficult will be done worse.

The evidence from our present results does not permit us to choose any one of these possible explanations to the exclusion of the others. The three types of possibility that we have put forward would seem, indeed, to be complementary rather than alternative, although the mechanisms postulated may be of different importance in different experimental situations. It would appear, therefore, that further work on difficulty in relation to transfer can profitably aim not at establishing the operation of one of these or other similar mechanisms to the exclusion of others, but at assessing quantitatively the extent of the operation of each.

We wish to record that Mr. C. B. Gibbs, working independently of us on a skilled task of a different kind, has noted (A.P.U. Report No. 113, 1949) transfer effects similar to those found by us. Also Mr. W. Adiseshiah and Mr. H. Kay, following Gibbs, have observed similar phenomena with other types of task. We gratefully acknowledge many valuable suggestions from them and from Mr. A. E. D. Schonfield during the preparation of this paper.

(*Received 22nd September*, 1949.)

BOOK REVIEW

The Biology of Mental Defect. By LIONEL S. PENROSE, M.A., M.D. With a Preface by Professor J. B. S. HALDANE, F.R.S. London: Sidgwick and Jackson, 1949. Pp. xiv + 285. With 7 plates. 21s.

There is Oliver Goldsmith's authority for saying that "were angels to write books they would not write folios"—most reviewers would agree. Professor Penrose, having written only a demy octavo, has not jeopardised his celestial prospects; and his book is exceptional in leaving an appetite for more. It has immediate relevance to at least four subjects—Psychology, Sociology, Medicine, and Human Genetics—and to review it fully would require a panel of specialists. Of his thirteen chapters, six deal with general topics—the historical background; definition, incidence, and classification of mental defect; principles of causation; and the methods of genetic analysis: six with the aetiology and natural history of specific forms of defect; and a final chapter with treatment. To psychologists the book's most obvious value lies in providing a comprehensive but not unwieldy account of mental deficiency, for which both teachers and students can be grateful. Its long-term importance, however, is probably to be found in the point of view, rather than in its practical utility in the class-room.

The word "Biology" in the title is one key to this point of view. Another is the remark that "the study" (of Human Genetics) "forms a descriptive and inductive science with some resemblance to Astronomy as opposed to Experimental Physics." Professor Penrose's attitude, in short, is that of the empirical (almost sceptical) biologist: and this carries with it certain modes of thought which are favourable to the conduct of research. It implies, in the first place, a descriptive, even taxonomic approach to the subject; only when the relevant variables have been defined in this way is exact statistical or experimental treatment possible. It also implies a regard for the whole individual, and for his setting. Human beings are the object of study and they are not to be set aside as the incidental repositories of subnormal I.Qs. Concepts such as Mental Defect, and Intelligence, in their common sense, have a considerable vagueness in definition—which is one reason why they are useful. They are no more measurable than is Health, and to identify them with what can be measured, e.g. performance on a test, is like regarding the clinical thermometer as a means of measuring Health. To borrow Ramsey's words: "the chief danger to our philosophy, apart from laziness and woolliness, is scholasticism, the essence of which is treating what is vague as if it were precise. . . ." Mental tests are seen as useful aids to clinical study, and no more; further, "each different mental test measures a different aspect of intelligence, just as measurements of stature, span, or girth measure different aspects of body size." Failure to perform adequately tasks of this kind is an important symptom, or symptom complex, but there are many other characteristics of defectives to be noted by the biologist. Testing the urine may, for instance, be more useful than testing the intelligence.

In conformity with this naturalistic attitude, is a refusal to be misled by the continuity of some metrical attribute into a belief that associated attributes or their causes must exhibit a similarly continuous variation. The belief that only a distinction of degree can be made between the normal and the mental defective, because both can be fitted into a continuous distribution of I.Qs. is no better founded in logic than the similar belief that a continuous distribution of scores on a test of "neuroticism," as between normals and neurotics, will allow us to dispense with the distinction between them. Professor Penrose in his consideration of the aetiology of defect is at pains to discover under what circumstances defectives may be regarded as normal variants and when abnormal determinants must be sought. A good example is to be found in the separation of cases of high and low grade defect—admittedly a classificatory rather than a fundamental distinction. Current teaching on this point too often relies on the fact that the distribution of Binet Test scores departs from normality at the bottom end. There are more low grade defectives than would be "expected"—therefore, the argument runs, some of them are aetiologically distinct from the others. The "expectation," however, presupposes that in test scores we have a direct linear measurement of the biological variable. A scalar distortion would equally well explain the observations. In summarising the qualitative differences which are actually found, Professor Penrose puts the distinction on much firmer ground.

I have considered points of methodology more explicitly than Professor Penrose treats them in his book and if I misrepresent him, I hope it may serve only to irritate him into

setting out more formally his own point of view. Nevertheless, I think all of this is implicit in his work, and largely taken for granted—as it probably would be taken for granted by most biologists, though perhaps not by most psychologists. *The Biology of Mental Defect* is not a polemical work, and often disapproval is conveyed quietly, but unmistakably by omission. Like Macaulay's conversation, after he returned from India, the work is "enlivened by occasional flashes of silence." A glance through the bibliography will show how many unedifying ideas we have been spared.

The writing is lucid and compact, and often marked by an attractively barbed form of understatement ("Goring expounded the view that . . . the criminal was, on the average . . . inferior to the man who arrested him. . . ." Nineteenth-century criminologists "initiated an enormous amount of measuring of prisoners.") In places the argument is perhaps over condensed, though this seems generally to be the result of a praiseworthy effort to keep the mathematics from proliferating unwholesomely. I doubt, for instance, whether anyone not already familiar with the subject would follow the derivation of the formulæ for consanguineous unions on pp. 100–101. Nevertheless, it is a considerable achievement to have written the chapter on Methods of Analysis in Human Genetics in terms which the non-mathematician should as a rule be able to follow. In the clinical sections he has avoided the gastronomic terminology popular with earlier descriptive writers: even the traditional "cherry-red spot" of the Tay-Sachs macula has been replaced by the more accurate if less succulent epithet "brownish-red." Illustrating these sections are photographs of a really hand-picked assortment of grotesques. The historical background is well described in an introductory chapter; and the opposing attitudes of religious bodies (the contrast lay originally between Eastern and Western Churches) are neatly summarised in Luther's assumption "that idiots were the illegitimate children of the devil," and the later view which christened them "*les enfants du bon Dieu*" (crétin, we are reminded, is a corruption of chrétien). Professor Penrose's own view inclines rather to the latter interpretation, and his chapter on treatment is devoted, encouragingly enough, to the things that can be done to help these children back into the world and not to the methods of assisting them out of it.

Since the book is likely to remain the standard work for a number of years, it is worth noting some points of detail, which might be corrected in a second impression. The distribution on p. 8 is due to Sir James Paget, who published it in the *St. Barthlomew's Hospital Gazette* in 1869; Galton borrowed it (with acknowledgments) 20 years later. The figure given for the proportion of the normal distribution outside 2 σ differs slightly but confusingly on pp. 26 and 27. The distinction between "incest" and "illegal unions" (pp. 57 and 58) is obscure, unless the Professor had in mind (which seems unlikely) the sin of spiritual incest. This, it is true, is not illegal under Common Law; otherwise the terms appear to be co-extensive (bigamous unions are excluded by the context). Affected males, in the notation of p. 164 should be hY not HY. And, finally, the spelling "Rohrschach," though pleasing, is unusual.

A. D. H.

EXPERIMENTAL PSYCHOLOGY GROUP

MEMBERS

BABINGTON SMITH, B., O.B.E., M.A. (Hon. Treasurer).	Institute of Experimental Psychology, 34, Banbury Road, Oxford.
CARPENTER, A., M.B., Ch.B.	Psychological Laboratory, Downing Place, Cambridge.
CHAMBERS, E. G., M.A.	Psychological Laboratory, Downing Place, Cambridge.
CRAWFORD, Miss A., M.A.	Department of Psychology, University of Liverpool, Liverpool.
DAVIS, D. RUSSELL, M.D. (Editor)	Psychological Laboratory, Downing Place, Cambridge.
DREW, G. C., B.Litt. (President)	Department of Philosophy and Psychology, The University, Bristol.
EYSENCK, H. J., M.A. Ph.D.	Institute of Psychiatry, The Maudsley Hospital, Denmark Hill, London, S.E.5.
FOSS, B. M., M.A.	Institute of Experimental Psychology, 34, Banbury Road, Oxford.
GRINDLEY, G. C., M.A., B.Sc.	Psychological Laboratory, Downing Place, Cambridge.
HARRIS, A. D., M.R.C.S., L.R.C.P.	Psychological Laboratory, Downing Place, Cambridge.
HALL, K. R. L., M.A., Ph.D.	Barrow Hospital, near Bristol.
HARPER, N., Ph.D.	National Institute for Research in Dairying, Shinfield, Reading.
HEIM, Miss A. W., M.A., Ph.D.	Psychological Laboratory, Downing Place, Cambridge.
HICK, W. E., M.D.	Psychological Laboratory, Downing Place, Cambridge.
HIMMELWEIT, Mrs. M. A., Ph.D.	Intsitute of Psychiatry, The Maudsley Hospital, Denmark Hill, London, S.E.5.
HOTOPH, N., M.A.	London School of Economics, Houghton Street, Aldwych, London, W.C.2.
JOYNSON, R. B., B.Litt.	Department of Psychology, University of Nottingham, Notts.
MACKWORTH, N. H., M.B., Ph.D.	Psychological Laboratory, Downing Place, Cambridge.
MACNEIL, Miss F. E., Ph.D.	Department of Psychology, University of Manchester.
OLDFIELD, R. C., M.A.	Institute of Experimental Psychology, 34, Banbury Road, Oxford.
PICKFORD, R. W., D.Litt., Ph.D.	Psychology Department, The University, Glasgow, W.2.
SEMEONOFF, B., Ph.D. (Committee)	Department of Psychology, The University of Edinburgh.
TRIST, E., O.B.E., M.A.	Institute of Human Relations, Tavistock House, 2, Beaumont Street, London, W.1.
VERNON, Miss M. D., M.A. (Committee)	Department of Psychology, The University, Reading.
VINCE, Miss M. A., M.A.	Psychological Laboratory, Downing Place, Cambridge.
WELFORD, A. T., M.A.	Psychological Laboratory, Downing Place, Cambridge.
WHITFIELD, J. W , M.A. (Committee)	Psychological Laboratory, Downing Place, Cambridge.
ZANGWILL, O. L., M.A. (Hon. Secretary)	Institute of Experimental Psychology, 34, Banbury Road, Oxford.

Visiting Foreign Members

ADISESHIAH, W., M.A.	Psychological Laboratory, Downing Place, Cambridge.
RUSSELL, R. W., Ph.D.	Bethlem Royal Hospital, Monks' Orchard, Eden Park, near Beckenham, Kent.